Deep Underground Neutrino Experiment (DUNE)

Far Detector Interim Design Report

Volume 2:

Chapter Breakout:

`far-detector-single-phase.tex`
`chapter-fdsp-daq.tex`

July 19, 2018

The DUNE Collaboration

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Contents

List of Figures

List of Tables

Chapter 1

Data Acquisition System

1.1 Data Acquisition (DAQ) System Overview

1.1.1 Introduction

The DUNE far detector (FD) data acquisition (DAQ) system must enable the readout, triggering, processing and distribution to permanent storage of data from all detector modules, which includes both their electrical time projection chamber (TPC) and optical photon detection system (PDS) signals. The final output data must retain, with very high efficiency and low bias, a record of all activity in the detector that pertains to the recognized physics goals of the DUNE experiment. The practical constraints of managing this output requires that the DAQ achieve these goals while reducing the input data volume by almost four orders of magnitude.

The current generation of liquid argon time-projection chamber (LArTPC) DAQs, such as used in ProtoDUNE and MicroBooNE, produce data spanning a fixed window of time that is chosen based on the acceptance of an external trigger. The DUNE DAQ faces several major challenges beyond those of the current generation. Foremost, it must accept data from about two orders of magnitude more channels and from that data it must form its own triggers. This self-triggering functionality requires immediate processing of the full-stream data from a large portion of all TPC channels with a throughput of approximately one terabyte per second per detector module. From this data stream, triggers must be raised based on two very different patterns of activity. The first is activity localized in a small region of one detector module, such as due to beam neutrino interactions or the passage of relatively rare cosmic-ray muons. This activity tends to correspond to a relatively large deposition of energy, around 100 MeV or more. The second pattern that must lead to a trigger is lower energy activity dispersed in both time and spatial extent of the detector module, such as due to a supernova neutrino burst (SNB).

The DAQ must also contend with a higher order of complexity compared to the current generation. The FD is not monolithic but ultimately will consist of four detector modules each of 10 kt fiducial

mass. Each module will implement somewhat different technologies and the inevitable asymmetries in the details of how data are read out from each must be absorbed by the unified DAQ at its front end. Further, each detector module is not monolithic but has at least one layer of divisions, here generically named detector units. For example, the single-phase (SP) detector module has anode plane assemblies (APAs) each providing data from a number of warm interface boards (WIBs) and the dual-phase (DP) detector module has charge readout (CRO) and light readout (LRO) units associated with specific electronics crates. In each detector module, there are on the order of 100 detector units (150 for SP and 245 for DP) and each unit has a channel count that is of the same order as that of an entire LArTPC detector of the current generation. The DUNE DAQ, composed of a cohesive collection of DAQ instances called DAQ partitions, must run on a subset of all possible detector units for each given detector module. Each instance effectively runs independently of all the others, however some instances indirectly communicate through the exchange of high-level trigger information. This allows, for example, each detector module to take data in isolation. It also allows for all detector modules to contribute to forming and accepting global SNB triggers, and to simultaneously run small portions – consisting of a few detector units – separately in order to debug problems, run calibrations or perform other activities while not interfering with nominal data taking in order to maintain high uptime.

Substantial computing hardware is required to provide the processing capability needed to identify such activity while keeping up with the rate of data. The nature of various technical, financial and physical constraints leads to the need for much of the computing hardware required for this processing to reside underground, near the detector modules. In such an environment, power, cooling, space, and access is far more costly than in typical data centers.

Past LArTPC and long-baseline (LBL) neutrino detectors have successfully demonstrated external triggering using information related to their beam. The DUNE FD DAQ will accept external information on recent times of Main Injector beam spills from Fermilab. This will assure triggering with high efficiency to capture activity pertaining to interactions from the produced neutrinos.

However, even if the DUNE experiment were interested only in neutrinos from beam spills, an external beam trigger alone would not be sufficient. Absent any other information, such a trigger must inevitably call for the readout of all possible data from the FD over at least one LArTPC drift time. This would lead to an annual data volume approaching an exabyte (10^{18} bytes), the vast majority of which would consist of just noise. This entire data volume would have to be saved to permanent storage and then processed offline in order to get to the signals.

DUNE's physics goals of course extend beyond beam-related interactions, including cosmic-ray muons, which provide an important source of detector calibration, and atmospheric neutrino interactions, which give a secondary source from which to measure neutrino properties. Taken together, recording their activity will dominate the data rate. The DAQ must also record data with sensitivity to rare interactions (both known and hypothetical) such as nucleon decay, other baryon number violating processes (such as neutron-antineutron oscillation), and interactions from the products of SNBs as well as possibly being able to observe isolated low-energy interactions from solar neutrinos and diffuse supernova neutrinos.

Some of these events, while rare in themselves, produce patterns of activity that can be mimicked by other higher-rate backgrounds, particularly in the case of SNBs. While the exact processes involved

in SNBs are not fully understood, it is expected that a prolonged period of activity of many tens of seconds will occur over which their neutrino interactions may be observed. Individually, these interactions will be of low energy (relative to that of beam neutrino interactions, for example), and will be spread over time and over the bulk of the detector modules. Because of their signature and their importance, special attention is required to first ascertain that a SNB may be occurring and to save as much data as possible over its duration.

Thus the DAQ must greatly reduce the full-stream of its input data while using the data itself to do so. It must do this efficiently both in terms of recording essentially all activity important to the physics goals of DUNE and in terms of a rate of data output that is manageable. To perform these primary duties the DAQ provides run control, configuration management, monitoring of both its processes and the general health of the data, and a user interface for these activities.

1.1.2 Design Considerations

The different detector modules vary in terms of their readout technology and schemes, timing systems, channel counts and data throughput and format. These aspects determine the nature of the digital data input to the DAQ. The design of the DAQ strives to contain the unique layers that adapt to the variation in the detector modules toward its front end in order to allow as many of its back end components to remain as identical across the detector modules as possible. In particular, the DAQ must present a unified interface to the ultimate consumer of its data, DUNE offline computing. It must also accept and process the data from a variety of other sources including the accelerator, various calibration systems (including laser, cold electronics (CE), photon detectors (PDs), and potentially others) as well as trigger sources external to DUNE. The modular nature of the DUNE FD implies that the DAQ instances running on each module must also exchange trigger information. In particular, exchanging module-local SNB trigger information will allow higher efficiency for this important physics. The DAQ must be optimized for the above while also retaining the flexibility to scale to handle risks such as excess noise, changes in high voltage (HV), cut network connectivity and other issues that could arise.

Currently, two major variations for the DUNE DAQ are under consideration. The eventual goal is to reduce this to a single high-level design which will service both SP and DP detector modules and be reasonably expected to support the third and forth modules to come. The first design, designated in this proposal as *nominal*, is illustrated in a high-level way in terms of its data and trigger flow in Figure 1.1. The second design, designated as the alternate, is similarly illustrated in Figure 1.2. The two variants differ largely at their FEs in terms of the order in which they buffer the data received from the detector module electronics and use it to form trigger primitives. They also differ in how they treat triggering and data flow due to a potential SNB. As their FEs are also sensitive to differences between the detector module electronics, this further variation for each general design is described below in Sections 1.2.2 and 1.2.3 for the detector module specific to this volume.

At this general high level, the two designs are outlined. For both, the diagrams are centered on one DAQ front-end fragment FE, which is a portion of the entire DAQ partition servicing a detector module that has one front-end computer (FEC) accepting about 10 to 20 Gbit/s of data

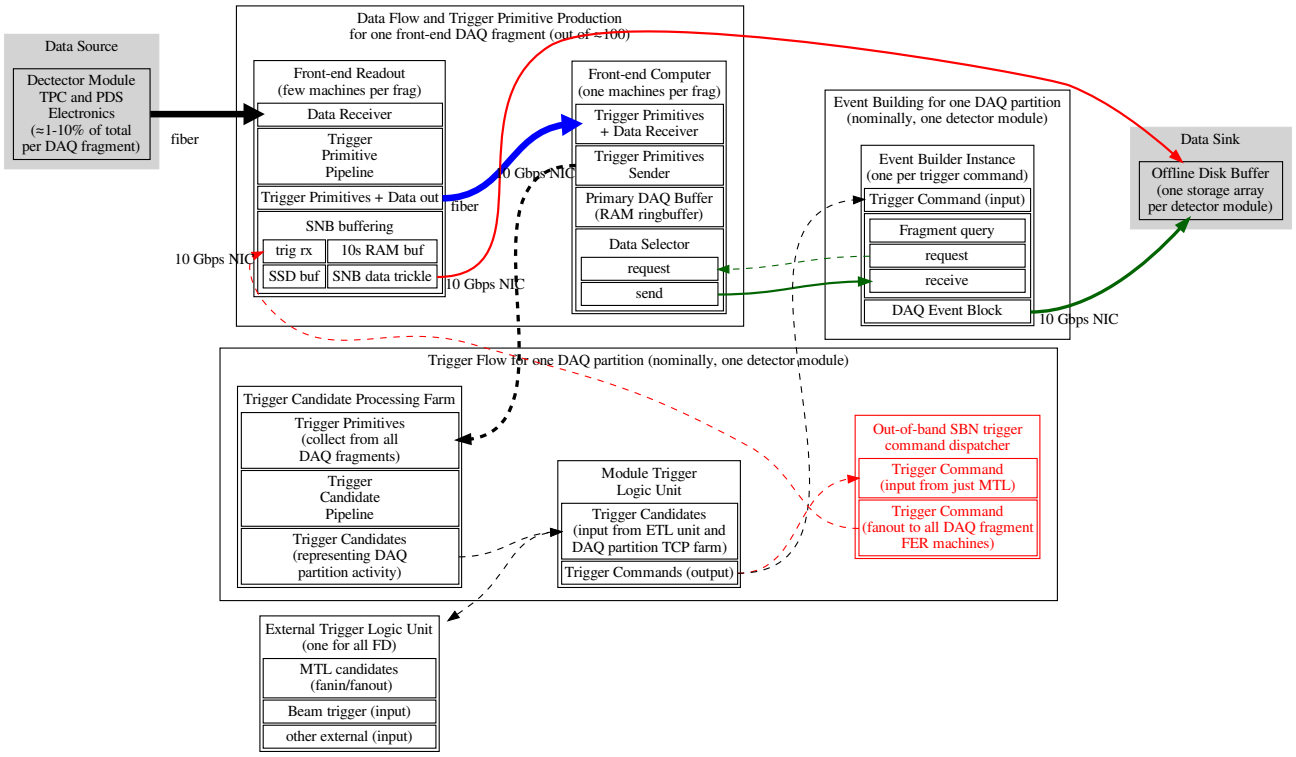

Figure 1.1: The high-level, *nominal* design for the DUNE FD DAQ in terms of data (solid) and trigger (dashed) flows between one DAQ front-end fragment FE and the trigger processing and event building back end for one DAQ partition. Line thickness indicates relative bandwidth requirements. Blue indicates where the full data flow for the DAQ front-end fragment is concentrated to one endpoint. Green indicates final output of normally triggered (non-SNB) data. Red indicates special handling of potential SNB. Each detector module has specialized implementation of some of these high level components, particularly toward the upstream FE as described in the text. The grayed boxes are not in the DAQ scope.

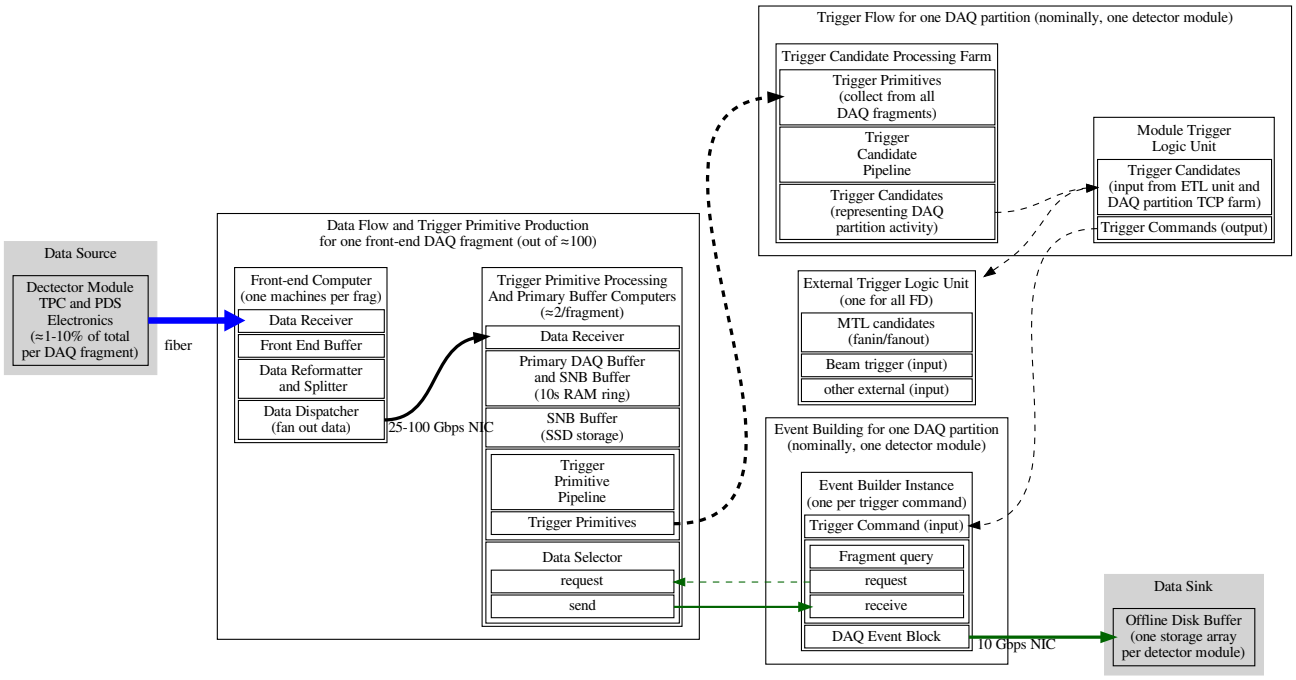

Figure 1.2: The high-level, alternate design for the DUNE FDFD DAQ in terms of data (solid) and trigger (dashed) flows between one DAQ front-end fragment FE and the trigger processing and event building back end for one DAQ partition. Line thickness indicates relative bandwidth requirements. Blue indicates where the full data flow for the DAQ front-end fragment is concentrated to one endpoint. Green indicates final output. Note, except for a longer readout, SNB is handled symmetric to normal data. Each detector module has specialized implementation of some of these high level components, particularly toward the upstream front-end as described in the text. The grayed boxes are not in the DAQ scope.

(uncompressed rate) from some integral number of detector units. Each of the participating DAQ front-end fragments do the following:

- Accept TPC and PDS data from the detector units associated with the DAQ front-end fragment.
- Produce and emit a stream of per-channel trigger primitives.
- Buffer the full data stream long enough for the trigger decision to complete (at least 10 s as driven by SNB requirements).
- Accept data selection requests and return corresponding data fragments.

All participating DAQ front-end fragments in the particular DAQ partition (i.e., the DAQ instance) servicing a portion of one detector module communicate with one trigger processing and event building system. The trigger processing system must:

- Receive the stream of per-channel trigger primitives from all DAQ front-end fragments.
- Correlate the primitives in time and spatially (across channels), and otherwise use them to form higher-level trigger candidates.
- Exchange trigger candidates with the external trigger logic (ETL).
- From them form trigger commands, each of which describes a portion of the data in time and a channel to be read out, such that no two trigger commands overlap.
- Dispatch these commands as required (in general to the event builder (EB)).

The event building system is responsible for performing the following actions:

- Accept a trigger command and allocate one EB instance to dispatch it.
- Interpret and execute the command by making data selection requests to referenced DAQ front-end fragments.
- Accept the returned data fragment from each DAQ front-end fragment and combine them into a DAQ event block.
- Write the result to the secondary DAQ buffer, which is the boundary shared with DUNE offline computing.

The nominal and the alternate DAQ designs differ largely in where the trigger primitive and SNB buffering exist. The *nominal* design places these functions in machines comprising a DAQ front-end readout (FER), which is upstream of the FEC. This then requires the SNB data and trigger handling to be different than that for normal (non-SNB) data. When a SNB trigger command is raised it is forwarded to the out-of-band trigger command dispatcher (OOB dispatcher) which sends

it down to the FERs. After the SNB data is dumped to solid-state disks (SSDs) it is "trickled" out via a path separate from the normal data to the secondary DAQ buffer. The *alternate* design, on the other hand, places these functions downstream of the FEC in trigger processing and data buffering nodes. The RAM of these nodes is used to provide the primary DAQ buffer for normal triggering as well as the deeper buffers needed for SNB. This design handles the SNB data somewhat symmetrically with normal data. When an EB makes a request for SNB data, it differs only in its duration, spanning tens of seconds of instead just a few milliseconds. The FE buffering nodes, instead of directly attempting to return the full SBN data immediately, streams it to local SSD storage. From that storage, the data is sent to the EB as low priority (i.e., also trickled out). Since the module trigger logic (MTL) ensures no overlapping commands, the buffer nodes may service subsequent requests from post-dump data that is in the RAM buffer. Since each trigger command is handled by an individual EB instance, the trickle proceeds asynchronously with respect to any subsequent trigger command handled by another EB instances.

Further description of these designs is given in Section 1.2.

The most critical requirements for the DUNE FD DAQ are summarized in Table 1.1.

Table 1.1: Important requirements on the DAQ system design

Requirement	Description
Scalability	The DUNE FD DAQ shall be capable of receiving and buffering the full raw data from all four detector modules
Zero deadtime	The DUNE FD DAQ shall operate without deadtime under *normal* operating conditions
Triggering	The DUNE FD DAQ shall provide full-detector triggering functionality as well as self-triggering functionality; the data selection shall maintain high efficiency to physics events while operating within a total bandwidth of 30 PB/year for all operating detector modules
Synchronization	The DUNE FD DAQ shall provide synchronization of different detector modules to within 1 μs, and of different subsystems within a module to within 10 ns

The input bandwidth and processing needs of the DAQ are expected to be dominated by the rate of data produced by the TPC system of each detector module. These rates vary between the modules and their estimations are summarized in Table 1.2.

The ultimate limit on the output data rate of the DUNE FD DAQ is expected to be provided by the available bandwidth to the tape, disk and processing capacity of Fermilab. An ample guideline has been established that places this limit at about 30 PB/year or 8 Gbit/s. Extrapolating to four detector modules, this requires a DAQ data reduction factor of almost four orders of magnitude. This is achieved through a simple self-triggered readout strategy.

An overestimate of the annual triggered but uncompressed data volume for one 10 kt detector module is summarized in Table 1.3. It assumes a very generous and simple trigger scheme whereby the data from the entire detector module is saved for a period longer than two drift times around

Table 1.2: The parameters governing the pre-trigger data rate from units of each detector module TPC CEs and the aggregate throughput into the FECs of the DAQ DAQ front-end fragments. Compression is an estimate and will be reduced if excess noise is introduced.

Parameter	single-phase	dual-phase
TPC unit	APA	CRO crate
Unit multiplicity	150	240
Channels per unit	2560 (800 collection)	640 (all collection)
ADC sampling	2 MHz	2.5 MHz
ADC resolution	12 bit	12 bit
Aggregate from CE	1440 GB/s	576 GB/s
Aggregate with compression	288 GB/s (5×)	58 GB/s (10×)

the trigger time. This essentially removes any selection bias at the cost of recording a substantial amount of data that will simply contain noise. Detailed trigger efficiency studies still remain to be performed. Initial understanding indicates that trigger efficiency should be near 100 % for localized energy depositions of at least 10 MeV. Sub-MeV signals can be ascertained from noise in existing LArTPCs so the effective trigger threshold may be even lower with high efficiency. Of course, data rates rise quickly when the threshold drops into the range of an MeV. Additional simulation and use of early data will be used to better optimize this threshold.

The data volume estimates also assume that any excess noise beyond what is expected due to intrinsic electronics noise will not lead to an increase in trigger rates. If, for example, excess noise occurs such that it frequently mimics more than about 10 MeV of localized ionization, this would lead to an increase in various types of triggers and subsequently more data. However, at the same time, these estimates do not take into account that some amount of lossless compression of the TPC data will be achieved. In the absence of excess noise it is expected that a compression factor of at least 5× can be achieved with the SP data and up to 10× may be achieved with the DP data, although the actual factor achieved will ultimately depend on the level of excess noise experienced in each detector module. Studies using data from the DUNE 35 ton prototype and early MicroBooNE running have shown that a compression factor of at least 4× can be expected even in the case of rather high levels of excess noise.

One category that will be particularly sensitive to excess noise is the trigger primitives. As discussed further in Section 1.2.3, their primary intended use is as transient objects produced and consumed locally and directly by the DAQ in the trigger decision process. However, as their production is expected to be dominated by ^{39}Ar decays (absent excess noise) they may carry information that proves very useful for calibration purposes. Future studies with simulation and with early data will determine the most feasible methods to exploit this data. These may include committing all or a portion to permanent storage or potentially developing processes that can summarize their data while still retaining information salient to calibration.

Finally, it is important to note that early data will be used to evaluate other selection criteria. It is expected that efficient and bias-free selections can be developed and validated that save a subset of the entire detector module for any given trigger type. For example, a cosmic-muon trigger command for a SP module will indicate which anode plane assemblies contributed to its

Table 1.3: Anticipated annual, uncompressed data rates for a single SP module. The rates for normal (non-SNB triggers) assume a readout window of 5.4 ms. For planning purposes these rates are assumed to apply to a DP module as well, which has a longer readout time but fewer channels. In reality, application of lossless compression is expected to provide as much as a $5\times$ reduction in data volume for the SP module and as much as $10\times$ for the DP module.

Event Type	Data Volume PB/year	Assumptions
Beam interactions	0.03	800 beam and 800 dirt muons; 10 MeV threshold in coincidence with beam time; include cosmics
Cosmics and atmospherics	10	10 MeV threshold, anti-coincident with beam time
Front-end calibration	0.2	Four calibration runs per year, 100 measurements per point
Radioactive source calibration	0.1	Source rate $\leq$10 Hz; single fragment readout; lossless readout
Laser calibration	0.2	1×10^6 total laser pulses, lossy readout
Supernova candidates	0.5	30 seconds full readout, average once per month
Random triggers	0.06	45 per day
Trigger primitives	$\leq$6	All three wire planes; 12 bits per primitive word; 4 primitive quantities; ^{39}Ar-dominated

formation (i.e., which ones had local ionization activity). This command can then direct reading out these anode plane assemblies, possibly also including their neighbors, while discarding the data from all other anode plane assemblies. This may reduce the estimated 10 PB/year for cosmics and atmospherics by an order of magnitude. A similar advanced scheme can be applied to the DP module by retaining data for the given readout window from only the subset of CRO crates (and again, potentially their nearest neighbors) that contributed to the formation of the given trigger.

1.1.3 Scope

The nominal scope of the DAQ system is illustrated in Figure 1.1 by the white boxes. It includes the continued procurement of materials for, and the fabrication, testing, delivery and installation of the following systems:

- FE readout (nominal design) or trigger farm (alternate design) hardware and firmware or software development for trigger primitive generation.
- FE computing for hosting of DAQ data receiver (DDR), DAQ primary buffer (primary buffer) and data selector.
- Back-end computing for hosting MTL, EB and the OOB dispatcher processes.
- External trigger logic and its host computing.
- Algorithms to generate trigger commands that perform data selection.
- Timing distribution system.
- DAQ data handling software including that for receiving and building events.
- The online monitoring (OM) of DAQ performance and data content.
- Run control software, configuration database, and user interface
- Rack infrastructure in the central utility cavern (CUC) for readout electronics, FE computing, timing distribution, and data selection.
- Rack infrastructure on surface at SURF for back-end computing.

1.2 DAQ Design

1.2.1 Overview

The design for the DAQ has been driven by finding a cost-effective solution that satisfies the requirements. Several design choices have been made and two major variations remain to be studied. From a hardware perspective, the DAQ design follows a standard HEP experiment design, with customized hardware at the upstream, feeding and funnelling (merging) and moving the data into computers. Once the data and triggering information are in computers, a considerable degree of flexibility is available; the processing proceeds with a pipelined sequence of software operations, involving both parallel processing on multi-core computers and switched networks. The flexibility allows the procurement of computers and networking to be done late in the delivery cycle of the DUNE detector modules, to benefit from increased capability of commercial devices and falling prices.

Since DUNE will operate over a number of decades, the DAQ has been designed with upgradability in mind. With the fall in cost of serial links, a guiding principle is to include enough output bandwidth to allow all the data to be passed downstream of the custom hardware. This allows the possibility for a future very-fast farm of computing elements to accommodate new ideas in how to collect the DUNE data. The high output bandwidth also gives a risk mitigation path in case the noise levels in a part of the detector are higher than specified and higher than tolerable by the baseline trigger decision mechanism; it will allow additional data processing infrastructure to be added (at additional cost).

Digital data will be collected from the TPC and PD readout electronics of the SP and DP detector modules. These categories of data sources are viewed as essentially four types of subdetectors within the DAQ and follow the same overall data collection scheme as shown for the nominal design in Figure 1.1 and for the alternate design in Figure 1.2. The readout is arranged to allow making a trigger decision in a hierarchical manner. Initial inputs are formed at the channel level, then combined at the detector unit level and again combined at the detector module level. In addition, the trigger decision process combines information at this level that may come from the other detector modules as well as information from sources external to the DAQ. This hierarchical structure in forming and consuming triggers allows safeguards to be developed so that any problems in one cavern or in one detector unit of one detector module need not overwhelm the entire DAQ. It also allows a SNB to be recorded in all operational parts of the detector while others may be down for calibration or maintenance.

Generally speaking, the DAQ consists of data flow and trigger flow. The trigger flow involved in self-triggering originates from processing a portion of the data flow. The trigger flow is then consumed back by the DAQ in order to govern what portion of the data flow is finally written out to permanent storage. The nominal and alternate designs differ in where in the data flow the trigger flow originates.

In both designs, a single DAQ front-end fragment associates an integral number of detector units with one front-end computer (FEC). This fragment forms one conceptual unit of the FE DAQ. The

processing on a FEC is kept minimal such that each has a throughput limited by I/O bandwidth. The recently released PCIe v4 doubles the bandwidth from the prior version and thus we assume that ≈20 GB/s throughput (out of a theoretical 32 GB/s max) can be achieved based on tests using PCIe v3. In principle then, this allows one FEC to accept the data from: two (if uncompressed) or ten (if 5× compressed) of the 150 SP anode plane assemblies, ten of the 240 DP CRO crates given their nominal 10× compression or the uncompressed data from all five DP LRO crates.

In the nominal design, the data enters the DAQ via the fragment's DAQ front-end readout (FER) component. In the SP the FER consists of eight reconfigurable computing elements (RCEs) and in the DP it consists of a number of Bump On Wire (BOW) computers, (see Section 1.2.2 in each respective detector module volume). The FER is responsible for accepting that data and from it producing channel level trigger primitives. It is also responsible for forwarding compressed data and the primitives to the DAQ data receiver (DDR) in the corresponding FEC. The FER is also responsible for supplying transient memory (RAM) and non-volatile buffer in the form of SSD sufficient for SNB triggering and readout. The DDR accepts the full data stream and transfers it to the DAQ primary buffer of its DAQ front-end fragment. There it is held awaiting a query from the event builder (EB). When the EB receives a trigger command it uses the included information to query all appropriate data selectors and from their returned data fragments an DAQ event block is built and written to file on the secondary DAQ buffer. From there the data becomes responsibility of the offline group to transfer to Fermilab for permanent storage and further processing.

In the alternate design, the data is accepted directly by the DAQ data receiver (DDR) in a FEC from the detector electronics for the particular detector module. The data then flows into the primary buffer and the portion required for forming trigger primitives is dispatched to the trigger computers of the fragment for the production of trigger primitives. Current SSD technology may allow SSD to be directly mounted to the FEC to provide for the SNB dump buffer. Another solution, which puts less pressure on write throughput, is to distribute the SSD for the SNB dumps to the trigger computers. In order to supply enough CPU for trigger primitive pipelines it is expected that at least two hosts per FEC will be needed. While their CPUs are busy finding trigger primitives, their I/O bandwidth will be relatively unused and thus they provide synergistic, cost-effective hosting for the SSDs.

Regardless of where the trigger primitives are produced in either the nominal or alternate design, they are further processed at the DAQ front-end fragment level to produce trigger candidates. At this level, they represent possible activity localized in time and by channel to a portion of the overall detector module. The trigger candidates emitted by all DAQ front-end fragments are sent to the module trigger logic (MTL) associated with the DAQ partition. There, they are time ordered and otherwise processed to form trigger commands. At this level they represent activity localized across the detector module and over some period of time.

The DAQ partition (or DAQ instance) just introduced is the cohesive collection of DAQ parts. One DAQ partition operates essentially independently from any other, and there is typically one per detector module. In some cases multiple DAQ partitions may operate simultaneously in a detector module, such as when some fraction of detector units are undergoing isolated testing or calibration.

Each trigger command is consumed by a single EB instance in order to query back to the DAQ

front-end fragments of its DAQ partition as described above. In addition, the MTL of one module is exchanging messages in the form of trigger candidates with the others. For example, one module may raise a local SNB trigger candidate and forward it to all other modules. Each module is also emitting candidates to sinks and accepting them from sources of external trigger information.

The exact implementation of some of these high-level functions, particularly those near the FE, depends on the particular detector module. The required specialization and in general, more implementation-level details are described in the following sections. Subsequent description proceeds toward the DAQ back end including processes handling dataflow, triggering, event building and data selection.

1.2.2 Front-end Readout and Buffering

Figure 1.3 illustrates the SP-specific DAQ front-end fragment specializing from the generic, nominal design illustrated in Figure 1.1. Starting from the left, it shows the fiber optic connectivity pattern between the four connectors of five SP WIBs associated with each APA and the elements of one SP DAQ front-end readout (FER). In total, the FER is associated with two anode plane assemblies, each of which is serviced by one ATCA Cluster On Board (COB) hosting four compute units called reconfigurable computing element (RCE). Each RCE provides field programmable gate array (FPGA), RAM and SSD resources. Its primary functions include:

- Receive data from WIBs,
- Produce trigger primitives from collection channels (see Section 1.2.3),
- Compress the data,
- Forward data and trigger primitives to the FEC,
- Buffer data in RAM,
- Stream data from RAM to SSDs on receipt of a *dump* trigger command (such as is raised by an SNB candidate),

The data and trigger primitives from the eight RCEs are aggregated to a single Front-End Link eXchange (FELIX) PCIe board residing in the FEC. This is done via 16 10 Gbit/s optical fibers. With no data compression performed in the RCEs the bandwidth of these fibers will be close to saturated. If excess noise is not greater than experienced by MicroBooNE, then a lossless compression factor of at least five may be expected. If achieved, the total throughput into FELIX from the two anode plane assemblies is expected to be about 4 GB/s

The firmware running on the FELIX FPGA transfers the data and trigger streams to the host system RAM. This type of transfer has been demonstrated by ATLAS with a PCIe v3 FELIX board at a throughput up to 10 GB/s. The next generation of FELIX based on PCIe v4 is expected to obtain about a factor of two improvement. The trigger primitives are combined across channels

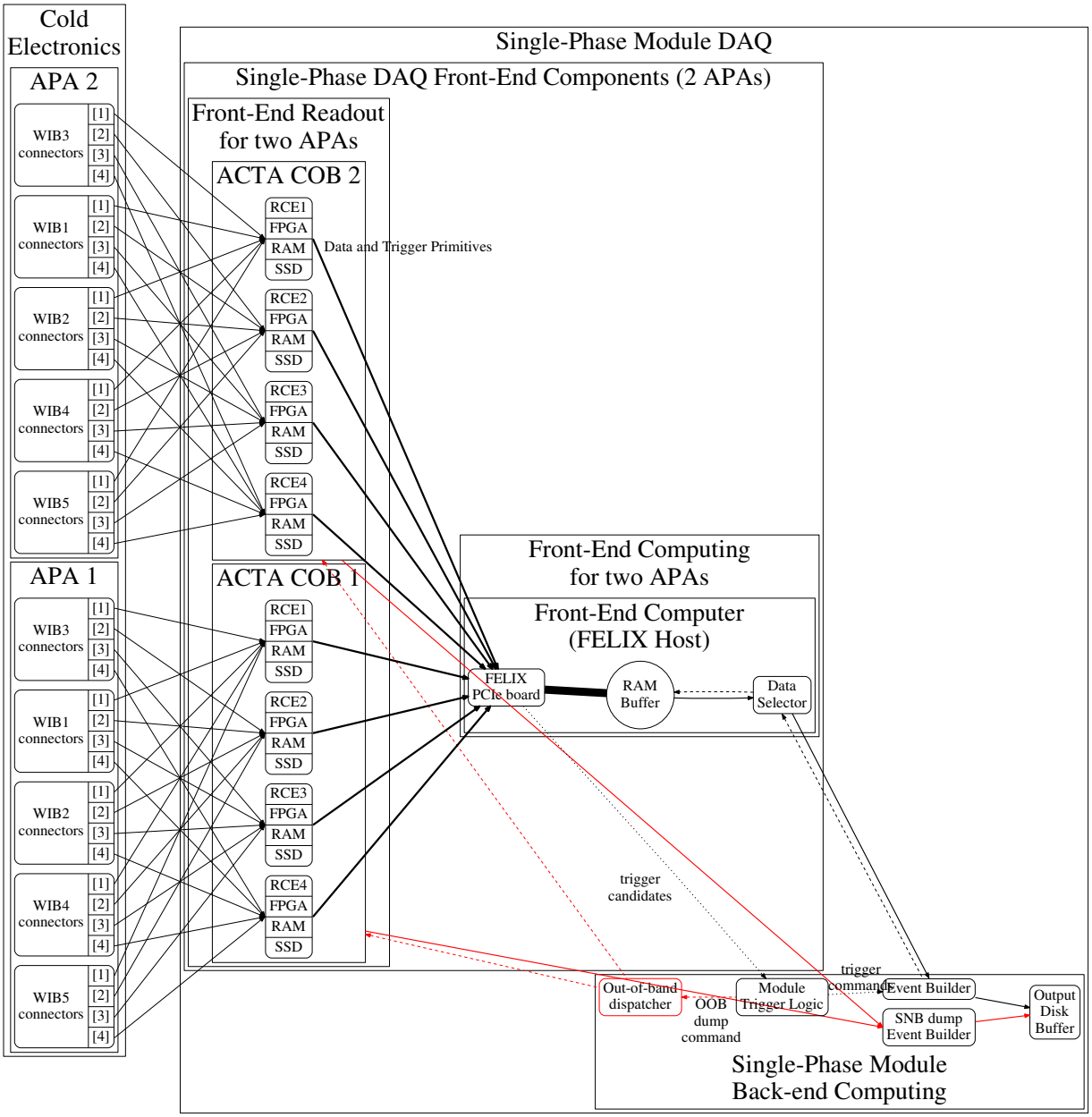

Figure 1.3: Illustration of data (solid arrows) and trigger (dashed) flow for one SP DAQ fragment (two APAs) in the nominal design. Black arrows indicate normal data and trigger flow and red indicate special flow for handling of a potential SNB.

to form trigger candidates that will be sent to the MTL. Meanwhile, the data stream streams into a primary buffer. This buffer will be sized sufficient to retain the full data stream for the period of time needed for a trigger decision to be made. As described above, that decision culminates in a trigger command that is sent to an EB. Based on its information, the EB makes a request to the data selector representing the DAQ front-end fragment, and the data selector replies with a data fragment build from the data available in the primary buffer.

An SNB trigger command is formed via the usual trigger hierarchy, as described in Section 1.2.3, and is consumed by the out-of-band trigger command dispatcher component. This component simply dispatches the command back down to the 600 RCEs in order to relieve the duty from the MTL, thus avoiding a source of trigger latency. This means an SNB trigger command is serviced differently than are all the other types of trigger commands. The RAM on board the RCEs is used to store the full data stream long enough for an SNB trigger command to be formed and distributed. It has been estimated that the rise time to detect an SNB in the SP detector module is about 1 s, so the RAM must be sized to buffer at least this much data. SNB models differ on the total duration over which significant neutrino interactions may be expected, as well as to their possible time profiles. Some allow for SNB neutrino interactions to occur for some time but at a rate not sufficient to rise above a trigger threshold determined by SNB backgrounds. It is assumed that an SNB dump should start about 10 s before the SNB trigger and should span a total 30 s. During the dump, all data is sent to both the SSD storage distributed among the RCEs. Data also continues to flow to the associated FER as during nominal running, which ensures that no dead time is suffered for all other non-SNB triggers. The multiplicities of RCEs and SSDs are such that uncompressed throughput of the SNB dump will just saturate current technology. Given a lossless compression factor of five the throughput to each SSD is expected to be 500 MB/s.

Given the infrequency of detectable SNBs the average SNB trigger rate is effectively governed only by a chosen threshold and by the rate of background from radiological decays, neutrons from cosmic-ray muons and fluctuations of noise, especially any coherent excess noise. The threshold must be tuned to maintain high efficiency for a broad class of SNB models while also not flooding the DAQ and potentially offline computing. This needs further study, but for the purposes of illustration a nominal false positive rate of one SNB dump per month is assumed. Uncompressed, this results in 540 TB/year. Each dump will take up 75 GB on an SSD, each of which is expected to provide 500 GB of storage; this is enough for at least six dumps. Again, lossless compression is expected to achieve a factor of five.

The SNB dumps are expected to remain on the SSD storage for some time in order to perform checks on the data to either rule out and delete the data or accept the candidate and migrate its data to permanent storage. This data migration is done out-of-band of the connection to the FELIX board using a local network connection to the ATCA crate. With the assumed average of one dump per month, if all were saved uncompressed it would require an average bandwidth aggregated over the entire SP module of just over 100 Mbit/s.

In the alternate SP DAQ design (not diagrammed), which corresponds to the generic Figure 1.2, the ten WIBs shown in Figure 1.3 are directly connected to one FELIX board via 10 Gbit/s fiber optic. The readout and buffering then follow the generic design.

1.2.3 Front-end Trigger Primitive Generation

Trigger primitives are generated inside the FE readout hardware associated with each APA from TPC data on a per-channel basis. They are sent along with the waveform data to the FE DAQ computing. In the alternate design, the data is directly sent from the anode plane assemblies to the FECs and the data are sent to the trigger processing computers. In both designs, the primitives from one DAQ front-end fragment are further processed to produce trigger candidates. As such they represent a localization of activity on the corresponding anode plane assemblies for a given period of time. These candidates are emitted to the MTL, which may consider candidates from other detector modules or external sources before generating trigger commands. Section 1.2.5 describes the selections involved in this triggering.

Only the 480 collection channels associated with each APA face are used for forming trigger primitives. Reasons for this limitation include the fact that collection channels:

- have higher signal to noise ratios compared to induction channels;
- are fully and independently sensitive to activity on their APA face;
- have unipolar signals that give direct approximate measures of ionization charge without the costly computation that would be needed to deconvolve the field response functions required for the the induction channels;
- can be easily divided into smaller, independent groups in order to better exploit parallel processing.

Figure 1.3 illustrates the connectivity between the four connectors on each of the five WIBs and the FE readout hardware. The data is received via eighty 1 Gbps fiber optical links by four RCEs in the Advanced Telecommunications Computing Architecture (ATCA) Cluster On Board (COB) system.

Due to the pattern of connectivity between WIBs and RCEs, each RCE receives the data from the collection channels that cover one contiguous half of one APA face. Each RCE has two primary functions. The first is transmission of all data as described in Section 1.2.4. The second is to produce trigger primitives from its portion of the collection channel data. The algorithms to produce the trigger primitives still require development but can be broadly described, as follows.

1. On a per channel basis, calculate a rolling baseline and spread level that characterizes recent noise behavior such that the result is effectively free of influence from actual ionization signals.
2. Locate contiguous runs of analog-to-digital converter (ADC) samples that are above a threshold defined in terms of the baseline and noise spread.
3. Emit their time bounds and total charge as a trigger primitive.

Each trigger primitive represents ionization activity localized (relatively) along the drift direction

by the times at which the signal crosses threshold and by two planes parallel to the collection wire and located midway between the wire and each of its neighboring wires. Depending on the threshold set, these trigger primitives may be numerous due to ^{39}Ar decays and noise fluctuations. If their rate cannot be sustained, the threshold may be raised or further processing may be done, still at the APA level, that considers more global information. This may be performed either in the RCEs or later in the FE computing hosts. In either case, the results are in the form of trigger candidates, which are sent to the MTL.

Sources of radio frequency (RF) emission inside the cryostat are minimized by design. Any residual RF is expected to be picked up coherently across some group of channels. Depending on its intensity, additional processing of the collection waveforms must be employed to mitigate this coherent noise and this must occur before the data is sent for trigger primitive production. If the required mitigation algorithms outgrow the nominally specified RCE FPGA it is possible to double the number of COBs per APA, which would require a redistribution of fibers. Alternatively, or in addition, the higher number of trigger primitives produced as a result of excess noise can be passed along for further processing in the FE computing. This would require reprocessing the raw waveform data.

1.2.4 Dataflow, Trigger and Event Builder

In the general data and trigger flow diagrams for the nominal (Figure 1.1) and alternate (Figure 1.2) designs, the dataflow, trigger and event builder functions take as input data from the detector module electronics and culminate in files deposited to the secondary DAQ buffer for transfer to permanent storage by offline computing processes. The continuous, uncompressed data rate of the input from one detector module is on the order of 1 TB/s. The final output data rate, for all detector modules operating at any given time is approximately limited to 1 GB/s.

To accept this high data-inflow rate and to apply the substantial processing needed to achieve the required reduction factor, which is on the order of 1000, the DAQ follows a distributed design. The units of distribution for the front end of the DAQ must match up with natural units of the detector module providing the data. This unit is called the DAQ front-end fragment and each accepts input at a rate of about 10 to 20 GB/s. The exact choice maps to some integral number of physical detector module units (e.g., SP anode plane assemblies or DP CROs and LROs).

As described in the previous sections, the nominal and alternate designs differ essentially in the order and manner in which the SNB buffering occurs and the trigger primitives are formed. The overall data flow, higher level triggering and building of *event* data blocks for final writing are conceptually very similar. This processing begins with the data being received by the FELIX PCIe board hosted in the FEC. The FELIX board performs a DMA transfer of the data into the primary buffer for the DAQ front-end fragment, which resides in the FEC host system RAM. This buffer is sized to hold ten seconds of data assuming the maximum uncompressed input rate associated with the fragment. While data is being written to the buffer, a delayed portion is also being read in order to dispatch it for various purposes. Any and all requests to further dispatch a subset of this data from the primary buffer must arrive within this buffer time. In the nominal design, the only dispatching will be from a request made by an EB (described more below) upon

receipt of a trigger command. In the alternate design, a suitable fraction of the data is also dispatched via high bandwidth (at least 25 to 50 Gbit/s simplex, less if data is compressed at this stage) network connections to a trigger farm so that trigger primitives may be formed. Whether the primitives are formed in this manner or extracted from the stream sent by the FER (as in the nominal design) these trigger primitives from one DAQ front-end fragment are collectively sent for further processing in order to be combined across channels and to then produce trigger candidates. These are finally combined for one detector module in the MTL. It is in the MTL where trigger candidates from additional sources are also considered, as described in section 1.2.5.

In both the nominal and alternate designs the dispatch of data initiated by normal (non-SNB) trigger commands is identical. This dispatch, commonly termed *event building* involves collection of data spanning an identical and continuous period of time from multiple primary buffers across the DAQ. As introduced above, each trigger command is consumed by an EB process. It uses fragment address information in the trigger command to query the data selector process representing each referenced DAQ front-end fragment and accepts the returned a data fragment. In the exceptional case that the delay of this request is so large that the primary buffer no longer contains the data, then an error return is supplied and recorded by the EB in place of the lost data. Such failures lead to indicators displayed by the detector operation monitoring system. The EB finally assembles all responses into a DAQ event block and writes it to file on the secondary DAQ buffer where it becomes the responsibility of DUNE offline computing.

The data selector and EB services are implemented using the general-purpose Fermilab data acquisition framework *art*DAQ for distributed data combination and processing. It is designed to exploit the parallelism that is possible with modern multi-core and networked computers, and has been used in ProtoDUNE and other experiments. The *art*DAQ framework is the principal architecture that will be used for the DUNE DAQ back-end computing. The authors of *art*DAQ have accommodated DUNE-specific requests for feature additions. Also, a number of libraries have been developed based on existing parts of *art*DAQ used to handle incoming data from data sources. It is likely that future DUNE extensions will be made by one of these two routes.

Unlike the dispatch of data initiated by a normal trigger commands, a command formed to indicate the possibility of a SNB is handled differently between the nominal and alternate designs. Such a command is interpreted to save all data from all channels for a rather extended time of 30 s starting from 10 s before the time associated with the trigger command. As no data selection is being performed, given the required bandwidth, special buffering to nonvolatile storage, in the form of SSD, is required. Today's technology supplies individual SSD in the M.2 expansion card form factor, which supports individual write speeds up to 2.5 GB/s. The two designs differ as to the location of and data source for these buffers.

In the nominal design, these SSDs reside in the FER as described in Section 1.2.2. In that location, due to larger granularity of computing units, the data rate into any one SSD is within the quoted write bandwidth. However, and as shown in Figure 1.1, the data and trigger flow for SNB in the nominal case takes a special path. Instead of an EB consuming the trigger command as described above, it is sent to the out-of-band trigger command dispatcher (OOB dispatcher), which dispatches it to each FER unit hosting an SSD. This component is used to immediately free up the MTL to continue to process normal triggers. When the command is received, each host must begin to stream data from its local RAM, supplying at least 10 s of buffer to the SSD, and

continue until the full 30 s has elapsed. While it is performing this dump it must continue to form trigger primitives and pass them and the full data stream to the connected FEC.

In the alternate design the same primary buffer provides the 10 s of pre-trigger SNB buffering. As in the nominal case, it must rely on fast, local SSD storage to sink the dump. Current SSD technology allows four M.2 SSD devices to be hosted on a PCIe board. Initial benchmarks of this technology show that such a combination can achieve 7.5 GB/s write bandwidth, which is short of linear scaling. To support the maximum of 20 GB/s, three such boards would be required. The alternate design presents a synergy between the need to dump high-rate data and the need to provide CPU to form the trigger primitives. With current commodity computing hardware it is expected that each FEC will need to be augmented with about two computers in the trigger farm. These trigger processors will need to accept the entire DP and three-eighths of the SP data stream from their DAQ front-end fragment. If they instead accept the entire stream, they can also provide RAM buffering and split up the data rate, which must be sunk to SSD buffers.

In both designs, the data dumped to SSD may contain precious information about a potential SNB. It must be extracted from the buffer, processed and either discarded or saved to permanent storage. The requirements on these processes are not easy to determine. The average period between actual SNBs to which DUNE is sensitive is measured in decades. However, to maintain high efficiency for capturing such important physics, the thresholds will be placed as low as feasible, limited only by the ability to acquire, validate and (if validated) write out the data to permanent storage. Notwithstanding, the (largely false positive) SNB trigger rate is expected to be minuscule relative to normal triggers. Understanding the exact rate requires more study, including using early data, but for planning purposes it is simply assumed that one whole-detector data dump will occur per month on average. Using the SP module as an example, and choosing the nominal time span for the dump to be 30 s, about 45 PB of uncompressed data would result. In the nominal SP DAQ design, this dump would be spread over 600 SSD units leading to 75 GB per SSD per dump. Thus, typical SSDs offer storage to allow any given dump to be held for at least one half year before it must be purged to assure storage is available for subsequent dumps. If every dump were to be sent to permanent storage, it would represent a sustained 0.14 Gbit/s (per detector module), which is a small perturbation on the bandwidth supplied throughout the DAQ network. Saved to permanent storage this rate integrates to 0.5 PB/year, which while substantial, is a minor fraction of the total data budget. The size of each dump is still larger than is convenient to place into a single file, so the SNB event-building will likely differ from that for normal triggers in that the entire dump is not held in a single DAQ event block. Finally, it is important to qualify that these rates assume uncompressed data. At the cost of additional processing elements, lossless compression can be expected to reduce this data rate by 5 to 10 × or alternatively allow lower thresholds that lead to the same factor of more dumps. Additional study is required to optimize the costs against the expected increase in sensitivity.

1.2.5 Data Selection Algorithms

Data selection follows a hierarchical design. It begins with forming detector unit-level trigger candidates inside the DAQ front-end fragment FE computing using channel-level trigger primitives. These are then used to form detector module trigger commands in the MTL. When executed,

they lead to readout of a small subset of the total data. In addition, trigger candidates are provided to the MTL from external sources such as the ETL in order to indicate external events such as beam spills, or SNB candidates detected by the other detector modules. In addition to supplying triggers to SuperNova Early Warning System (SNEWS), triggers from SNEWS or other cosmological detector sources such as LIGO and VIRGO can be accepted in order to possibly record low-energy or dispersed activity that would not pass the self-triggering. The latency of arrival for these sources must be less than the nominal 10 s buffers used to capture low-level early SNB activity. A high-level trigger (HLT) may also be active within the MTL. The hierarchical approach is natural from a design standpoint and it allows for vertical slice testing and running multiple DAQ partitions simultaneously during commissioning of the system or when debugging of individual detector units is required.

As discussed in Sections 1.2.2 and 1.2.3, trigger primitives are generated in either in FERs (in the nominal design) or in trigger processing computers (in the alternate design). In both designs, and for both SP and DP detector modules, only data from TPC collection channels (three-eighths of SP and all of DP channels) feed the self-triggering, as their waveforms directly supply a measure of ionization activity without computationally costly signal processing. The trigger primitives contains summary information for each channel, such as the time of any threshold-crossing pulse, its integral charge, and time over threshold. A channel with an associated trigger primitive is said to be *hit* for the time spanned by the primitive. Trigger primitives from one detector unit are then further processed to produce a trigger candidate. The candidate represents a cluster of hits across time and channel, localized to the detector unit. The candidates from all DAQ front-end fragments are passed to the MTL.

The MTL arbitrates between various trigger types, determines trigger priority and ultimately the time range and detector coverage for a trigger command, which it emits back to the FECs. The MTL assures that no trigger commands are issued that overlap in time or in detector channel space. It also may employ a HLT to reduce or aggregate triggers into fewer trigger commands so as to optimize the subsequent readout. For example, aggregating many small readouts into fewer but larger ones may allow for more efficient processing. This can be particularly important during periods of high-rate activity due to e.g., various backgrounds or instrumental effects.

When activity leads to the formation of a trigger command this command is sent down to the FECs instructing which slice of time of its buffered data should be saved. The trigger command information is saved along with this data. At the start of DUNE data taking, it is anticipated that for any given single-interaction trigger (a cosmic-ray track, for example), waveforms from all channels in the detector module will be recorded over a one readout window (nominally, 5.4 ms for SP and 16.4 ms for DP, chosen to be two drift times plus an extra 20 %).

Such an approach is clearly very generous in terms of the amount of data saved, but it ensures that associated low-energy physics (such as captures of neutrons produced by neutrino interactions or cosmic rays) are recorded without any need to fine-tune detector unit-level triggering, and does not depend on the noise environment across detector units. In addition, the wide readout window ensures that the data of all associated activity is recorded. As generous as it is, it is estimated that this readout window will not produce an unmanageable volume of data. As shown in Table 1.3, the uncompressed selected data from the SP module will fill about half of the nominal annual data budget. The longer DP drift and its fewer channels will give approximately the same data

rate. However, once a modest amount of lossless compression is applied, the nominal data budget can be met. Early running will allow experience to be gained and more advanced data selection algorithms to be validated allowing the DAQ to discard the many data fragments in each trigger consistent with just electronics noise. This has the potential for a reduction of at least another factor of ten.

Other trigger streams – calibrations, random triggers, and prescales of various trigger thresholds – are also generated at the detector module level, and filtering and compression can be applied based upon the trigger stream. For example, a large fraction of random triggers may have zero-suppression (ZS) applied to their waveforms, reducing the data volume substantially, as the dominant data source for these will be ^{39}Ar events. Additional signal-processing can also be done on particular trigger streams if needed and if the processing is available, such as fast analyses of calibration data.

At the detector module level, a decision can also be made on whether a series of interactions is consistent with an SNB. If the number of detector unit-level, low-energy trigger candidates exceeds a threshold for the number of such events in a given time, a trigger command is sent from the MTL back to the FERs, which store up to 10 s of full waveform data. That data is then streamed to non-volatile storage to allow for subsequent analysis by the SNB working group, perhaps as an automated process. If not rejected, it is sent out of the DAQ to permanent offline storage.

In addition, the MTL passes trigger candidates up to a detector-wide ETL, which among other functions, can decide whether, integrated across all modules, enough detector units have detected interactions to qualify as an SNB, even if within a particular module the threshold is not exceeded. Trigger candidates from the ETL are passed to the MTL for dispatch to the FECs (or FERs in the case of SNB dump commands in the nominal design). That is, to the MTL, an external trigger candidate looks like just one more *external* trigger input.

Detector unit level trigger candidates are generated within the context of one DAQ front-end fragment, specifically in each FEC. The trigger decision is based on the number of nearby channels hit in a given fragment within a time window (roughly 100 µs), the total charge collected in these adjacent channels, and possibly the union of time-over-threshold for the trigger primitives in the collection plane. Studies show that even for low-energy events (roughly 10 MeV to 20 MeV) the reduction in radiological backgrounds is extremely high with such criteria. The highest-rate background, ^{39}Ar, which has an overall rate of 10 MBq within a 10 kt volume of argon, has an endpoint of 500 keV and requires significant pileup in both space and time to get near a 10 MeV threshold. One important background source is ^{42}Ar, which has a 3.5 MeV endpoint and an overall rate of 1 kBq. ^{222}Rn decays via a 5.5 MeV kinetic energy α and is also an important source of background. The radon decays to ^{218}Po, which within a few minutes leads to a 6 MeV kinetic energy α, and ultimately to a ^{214}Bi daughter (many minutes later), which has a β decay with its endpoint near 3.5 MeV kinetic energy. The α ranges are short, resulting in charge being collected on one or two anode wires at most, but the charge deposit can be large, and therefore the charge threshold must be well above the α deposits plus any pileup from ^{39}Ar and noise.

At the level of one detector unit, two kinds of local trigger candidates can be generated. One is a high-energy trigger that indicates local ionization activity corresponding to more than than 10 MeV. The per-channel thresholds on total charge and time-over-threshold will be optimized to

achieve at least 50 % efficiency at this energy threshold, with efficiency increasing to 100 % via a turn-on curve that ensures at least 90 % efficiency at 20 MeV. The second type of trigger candidate generated is for low-energy events between 5 MeV and 10 MeV. In isolation, these candidates do not lead to formation of a trigger command. Rather, at the detector module level they are combined, time ordered and their aggregate rate compared against a threshold based on fluctuations due to noise and backgrounds in order to form an SNB trigger command.

The MTL takes as input trigger candidates (both low-energy and high-energy) from the participating DAQ front-end fragments, as well as external trigger candidate sources, such as the ETL, which includes global, detector-wide triggers, external trigger sources such as SNEWS, and information about the time of a Fermilab beam spill. The MTL also generates trigger commands for internal consumption, such as random triggers and calibration triggers (for example, telling a laser system to fire at a prescribed time). The MTL can also generate trigger commands from a prescaled fraction of trigger types that otherwise do not generate such commands on their own. For example, a prescaled fraction of single, low-energy trigger commands could be allowed to generate a trigger command, even though those candidates normally only result in a trigger command when aggregated (i.e., as they would be for an SNB).

The MTL is also responsible for checking candidate triggers against the current run control (RC) trigger mask: in some runs, for example, we may decide that only random triggers are accepted, or that certain trigger candidates streams should not be considered because their DAQ front-end fragments have been producing unreasonably large rates in the recent past (such as may be due to noise spikes, flaky hardware or buggy software). In addition, the MTL counts low-energy trigger candidates, and based upon their number and distribution over a long time interval (e.g., 10 s), decides to generate an SNB trigger command. The trigger logic will be optimized to record the data due to at least 90 % of all Milky Way supernovae, and studies of simple low-energy trigger criteria show that a much higher efficiency can likely be achieved.

The HLT can also be applied at this level, particularly if there are unexpectedly higher rates from instrumental or low-energy backgrounds that require some level of reconstruction or pattern recognition. An HLT might also allow for efficiently triggering on lower-energy single interactions, or allow for better sensitivity for supernovae originating outside the Milky Way galaxy, by employing a weighting scheme to individual trigger candidates – higher-energy trigger candidates receiving higher weights. Thus, for example, two trigger candidates consistent with 10 MeV interactions in 10 s might be enough to create a SNB candidate trigger, while a hundred 5 MeV trigger candiates in 10 s might not. Lastly, the HLT can allow for dynamic thresholding; for example, if a trigger appears to be due to a cosmic-ray muon, the threshold for single interactions can be lowered (and possibly prescaled) for a short time after that to identify spallation products. In addition, the HLT could allow for a dynamic threshold after a SNB, to extend sensitivity beyond the 10 s SNB readout window, while not increasing the data volume associated with SNB candidates linearly.

All low-energy trigger candidates are also passed upwards to the ETL so that they may be integrated across all 10 kt detector modules in order to determine that a SNB may be occurring. This approach increases the sensitivity to trigger on SNBs by a factor of four (for 40 kt), thus extending the burst sensitivity to a distance twice as far as for a single 10 kt detector module.

The MTL is also responsible for including in the trigger command a global timestamp built from

its input trigger candidates, and information on what type of trigger was created. Information on trigger candidates is also kept, whether or not they contribute to the formation of a trigger command. As described above, the readout window for nominal trigger commands (those other than for SNB candidates) is somewhat more than two times the maximum drift time. Further, a nominal readout spans all channels in a detector module. The MTL is also responsible for sending the trigger commands that tell the FERs to stream all data from the past 10 s and for a total of 30 s in hopes to catch SNBs. This command may be produced based on trigger candidates from inside the MTL itself or it may be produced based on an external SNB trigger candidate passed to the MTL by the ETL.

1.2.6 Timing and Synchronization

All components of the SP module are synchronized to a common clock. In order make full use of the information from the PDS, the common clock must be aligned within a single detector unit with an accuracy of O 1ns. In order to form a common trigger for SNB between detector modules, the timing between them must be aligned with an accuracy of O 1ms. However, a tighter constraint is the need to calibrate the common clock to universal time (derived from GPS) in order to adjust the data selection algorithm inside an accelerator spill, which requires an absolute accuracy of O 1µs.

SP and DP detector modules use different timing systems, driven by the different technical requirements and development history of the two technologies. A SP module has many more timing end points than a DP module and many of the end points are simpler than the end points in the DP, for example a WIB versus Micro Telecommunications Computing Architecture (μTCA) crate. Both systems have been sucessfully prototyped.

The DUNE SP module uses a development of the ProtoDUNE-SP timing system. Synchronization messages are transmitted over a serial data stream with the clock embedded in the data. The format is described in DUNE DocDB-1651 [2]. Figure 1.4 shows the overall arrangement of components within the SP Timing System (SPTS). A stable master clock, disciplined with a 10 MHz reference is used in the SPTS. A one-pulse-per-second signal (1PPS signal) is also received by the system and is time-stamped onto a counter clocked by the SPTS master clock, however the periodic synchronization messages distributed to the SP detector module are an exact number of clock cycles apart even if there is jitter in the 1PPS signal.

The GPS signal is encoded onto optical fiber and transmitted to the CUC, where it is converted back to an RF signal on coaxial cable and used as the input to a GPS displined oscillator. The oscillator module also houses a IEEE 1588 (PTP) grandmaster and an NTP server. The PTP grandmaster provides a timing signal for the DP White Rabbit (WR) timing network. The NTP server provides an absolute time for the 1PPS signal. The SPTS relates its time counter onto GPS time by timestamping the 1PPS signal onto the SPTS time counter and reading the time in software from the NTP server.

The latency from the GPS antenna on the surface to the GPS receiver in the CUC will be measured by optical time domain reflectometry at installation. Given the modest absolute time accuracy

required (sufficient to select data within an accelerator spill) dynamic monitoring of this delay is not required.

The WR synchronization signals from the DP detector module are time-stamped onto the SPTS clock domain and the SPTS synchronization signals are time stamped onto the DP clock domain. This allows the timing in the SP and DP detector modules to be aligned. A similar scheme is used to relate the ProtoDUNE-SP timing domain to the beam instrumentation WR time domain.

In order to provide redundancy, and also the ability to easily detect issues with the timing path, two independent GPS systems are used. One with an antenna at the head of the Yates Shaft, the other with an antenna at the head of the Ross Shaft. The two independent timing paths are brought together in the same rack in the CUC. Using 1:2 fiber splitters one SPTS unit can be left as a hot spare while the other is active. This also allows testing of new firmware and software during comissioning without the risk of losing the SPTS if a bug is introduced.

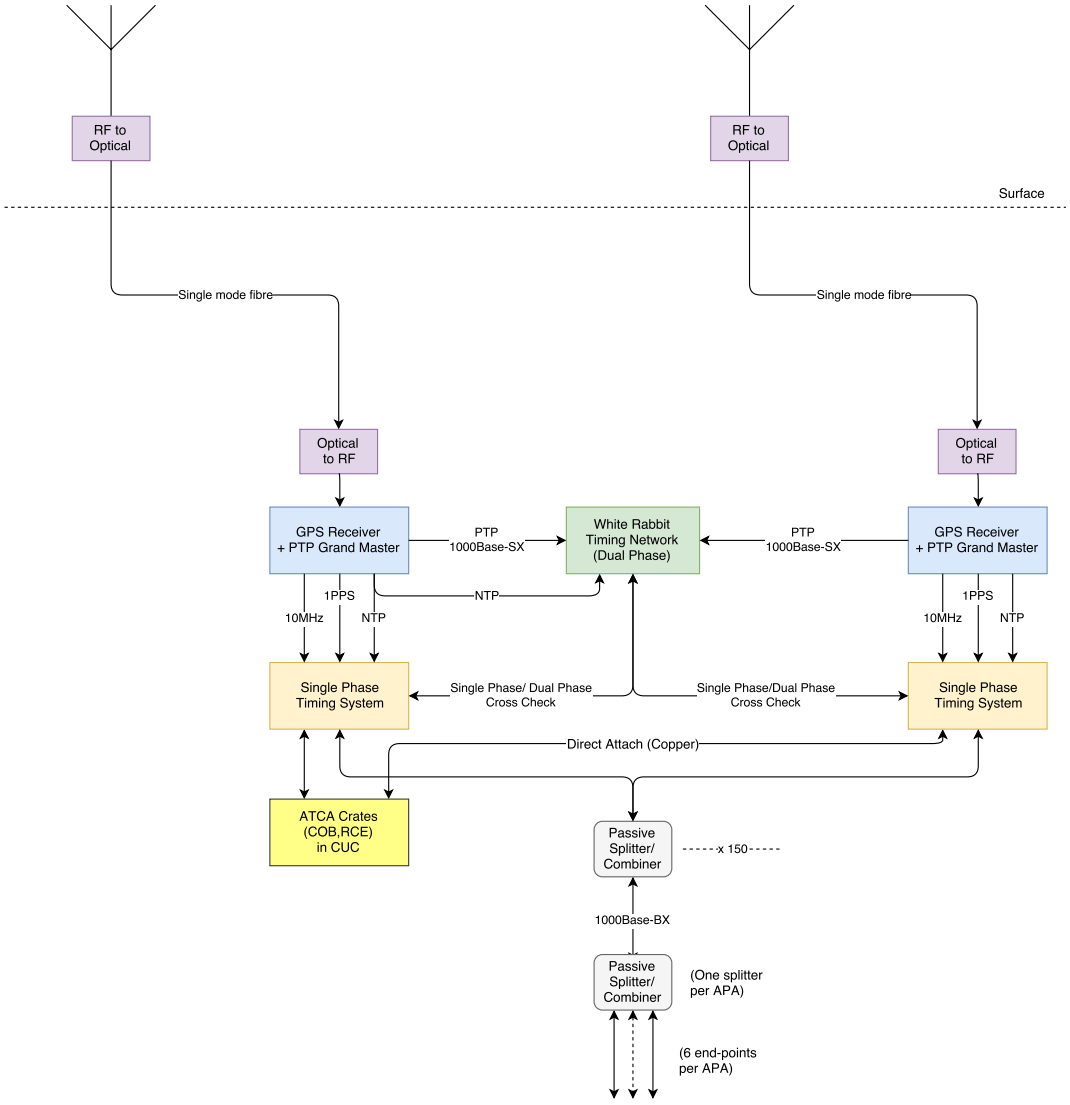

Figure 1.4: Illustration of the components in the DUNE timing system.

All the custom electronic components for the SPTS are contained in two μTCA shelves. At any

one time one is active and the other is a hot spare. The 10 MHz reference clock and the 1PPS signal are received by a single width advanced mezzanine card (AMC) at the center of the μTCA shelf. This master timing AMC produces the SPTS signals and encodes them onto a serial data stream. This serial datastream is distributed over a standard star-point backplane to the fanout AMCs, which each drive the signal onto up to 13 SFP cages. The SFP cages are either occupied by 1000Base-BX SFPs, each of which connects to a fiber running to an APA, or to a Direct Attach cable which connects to systems elsewhere in the CUC, i.e., the RCE crates and the data selection system. This arrangement is shown in Figure 1.5

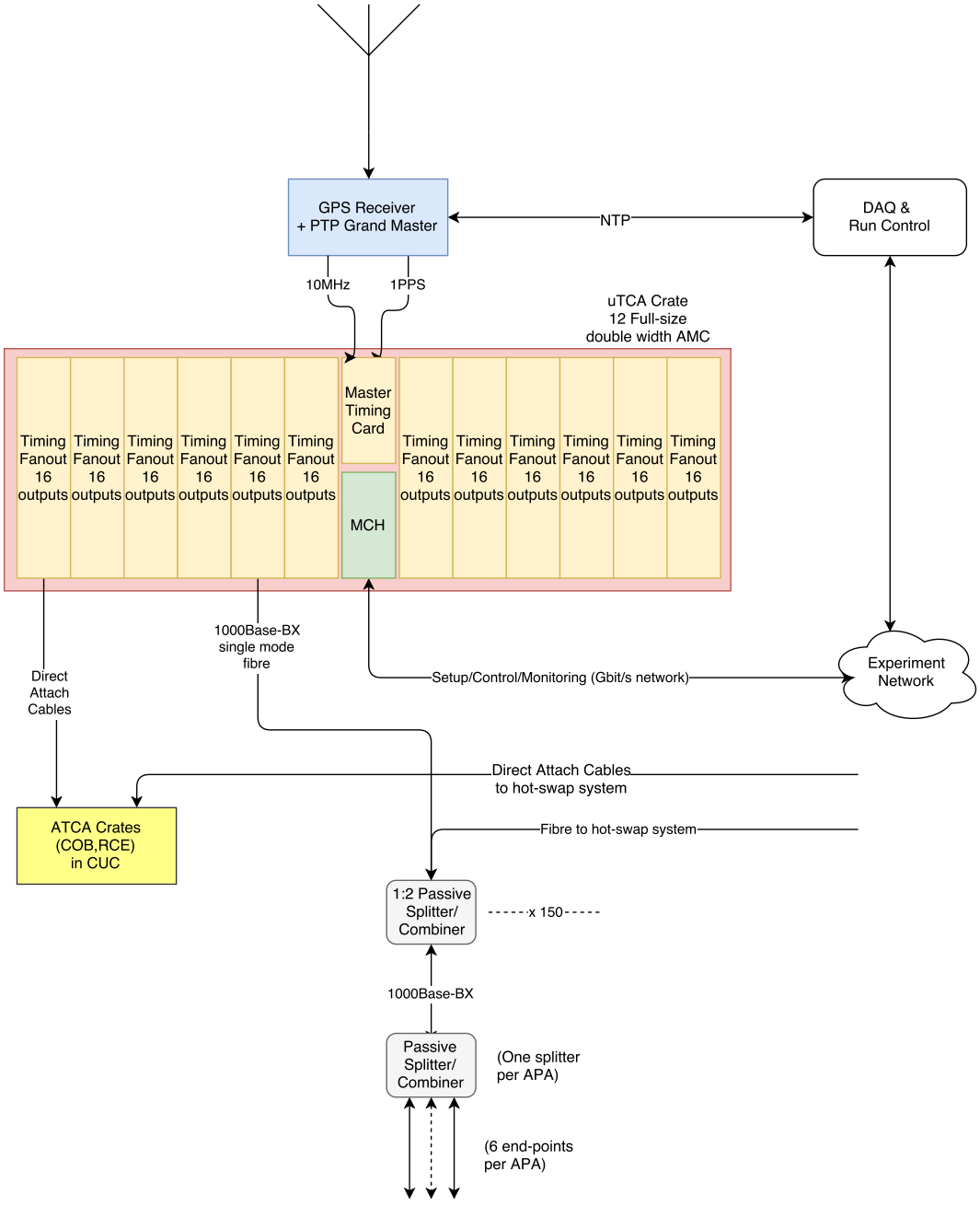

Figure 1.5: Illustration of the components in the SP timing system.

1.2.6.1 Beam timing

The neutrino beam is produced at the Fermilab accelerator complex in spills of 10 µs duration. A spill location system (SLS) at the Far Detector site will locate the time periods in the data when beam could be present, based on network packets received from Fermilab containing predictions of the GPS-time of spills soon to occur or absolute time stamps of recent spills. Experience from MINOS and NOνA shows that this can provide beam triggering with high reliability with some small fraction of late or dropped packets. To improve reliability further, the system outlined here contains an extra layer of redundancy in the prediction process. Several stages of prediction based on recent spill behavior will be applied, aiming for an accuracy of better than 10% of a readout time (sub-ms) in time for the data to be selected from the DAQ buffers. Ultimately, an offline database will match the actual time of the spill with the data, thus removing any reliance on real-time network transfer for this crucial stage of the oscillation measurements. The network transfer of spill-timing information is simply to ensure that a correctly located and sufficiently wide window of data is considered as beam data. This system is not required, and is not designed to provide signals accurate enough to measure neutrino time-of-flight.

The precision to which the spill time can be predicted at Fermilab improves as the acceleration process of the protons producing the spill in question advances. The spills currently occur at intervals of 1.3 s; the system will be designed to work with any interval, and to be adaptable in case the sequence described here changes. For redundancy, three packets will be sent to the far detector for each spill. The first is approximately 1.6 s before the spill-time, which is at the point where a 15 Hz booster cycle is selected; from this point on, there will be a fixed number of booster cycles until the neutrinos and the time is subject to a few ms of jitter. The second is about 0.7 s before the spill, at the point where the main injector acceleration is no longer coupled to the booster timing; this is governed by a crystal oscillator and so has a few µs of jitter. The third will be at the so called '`$74`' signal generated before the beamline kicker magnet fires to direct the protons at the LBNF target; this doesn't improve the timing at the Far Detector much, but serves as a cross check for missing packets. This system is enhanced compared to that of MINOS-NOνA, which only use a third of the above timing signals. The reason for the larger uncertainty in the time interval from 1.6 s to 0.7 s is that the booster cycle time is synchronized to the electricity supply company's 60 Hz, which has a variation of about 1%.

Arrival-time monitoring information from a year of MINOS data-taking was analyzed, and it was found that 97% of packets arrived within 100 ms of being sent and 99.88% within 300 ms.

The SLS will therefore have estimators of the GPS-times of future spills, and recent spills with associated data contained in the primary buffers. These estimators will improve in precision as more packets arrive. The DAQ will use data in a wider window than usual, if, at the time the trigger decision has to be made, the precision is lower due to missing or late packets. From the MINOS monitoting analysis, this expected to be very rare.

1.2.7 Computing and Network Infrastructure

The computing and network infrastructure that will be used in each of the four detector modules is similar, if not identical. It supports the buffering, data selection, event building, and data flow functionality described above, and it includes computing elements that consist of servers that:

- buffer the data until a trigger decision is received;
- host the software processes that build the data fragments from the relevant parts of the detector into complete events;
- host the processes that make the trigger decision;
- host the data logging processes and the disk buffer where the data is written;
- host the real-time data quality monitoring processing;
- host the control and monitoring processes.

The network infrastructure that connects these computers has the following components:

- subnets for transferring triggered data from the buffer nodes to the event builder nodes; these need to connect underground and above-ground computers;
- a control and monitoring subnet that connects all computers in the DAQ system and all FE electronics that support Ethernet communication; this sub-network must connect to underground and above-ground computers;
- a subnet for transferring complete events from the event builder servers to the storage servers; this subnet is completely above-ground.

1.2.8 Run Control and Monitoring

The online software constitutes the backbone of the DUNE DAQ system and provides control, configuration and monitoring of the data taking in a uniform way. It can be subdivided logically into four subsystems: the run control, the management of the DAQ and detector module electronics configuration, the monitoring, and the non-physics data archival. Each of these subsystems has a distinct function, but their implementation will share underlying technologies and tools.

In contrast to experiments in which data taking sessions, i.e., runs, are naturally subdivided into time slots by external conditions (e.g., a collider fill, a beam extraction period), the DUNE experiment aims to take data continuously. Therefore, a classic run control with a coherent state machine and a predefined and concurrently configured number of active detector and DAQ elements does not seem adequate.

The DUNE online software is thus structured according to the architecture principle of loose coupling: each component has as little knowledge as possible of other components. While the granularity of the back-end DAQ components may match the individual software processes, for the front-end DAQ a minimum granularity must be defined, balancing fault tolerance and recovery capability against the requirement of data consistency. The smallest independent component is called a DAQ front-end fragment, which is made up of the detector units associated with a single front-end computer. In the nominal design, this corresponds to two SP anode plane assemblies and about ten DP CRO crates.

The concept of a *run* represents a period of time in which the same FE elements are active or the same data selection criteria are in effect (possibly with maximum lengths for offline processing reasons). More than just orchestrating data taking, the run control provides the mechanisms allowing DAQ applications to publish their availability, subscribe to information, and exchange messages. In addition, the online software provides a configuration service for DAQ elements to store their settings and a conditions archive, keeping track of varying detector electronics settings and status.

Another important aspect of the online software is the monitoring service. Monitoring can be subdivided into two main domains: the monitoring of the data taking operations (rates, number of data fragments in flight, error flags, application logs, network bandwidth, computing and network infrastructure) and the monitoring of the physics data. Both are essential to the success of the experiment and must be designed and integrated into the DAQ system from the start. In particular, for such a large and distributed system, the information sharing and archival system is very important, as are scalable and easily accessible data visualization tools, which will evolve during the lifetime of the experiment. The online software provides the glue that holds the DAQ applications together and enables data taking. Its architecture guides the approach to DAQ application design and also shapes the view that the operators will have of the experiment.

1.3 Interfaces

1.3.1 TPC Electronics

Details about the interfaces between the DAQ and the TPC electronics are documented for the SP detector modules in [3].

In the case of the SP module, data from the CE front-end mother boards (FEMBs) are 8b10 encoded and sent to the WIBs on copper cables at a bit rate of 1.28 Gbit/s. There are two options being considered for the WIBs. In one, the data are simply converted to optical signals and transmitted to the DAQ in the CUC on 1.28 Gbit/s optical links with a total of 80 fibres per APA. In the second option, the WIBs aggregate the data onto links running at ≈10 Gbit/s before transmission to the DAQ, with a total of ten fibres per APA. In both cases the data are received on rear transition modules connected to the COB ATCA boards (see Section 1.2.2).

1.3.2 PD Electronics

Details about the interfaces between the DAQ and SP photon detection system are documented in [4].

For the SP PDS the S/N ratio of the silicon photomultiplier (SiPM) signals is high enough to allow zero-suppression to be safely applied to the data. This reduces the data flow so that a bandwidth of 8 Gbit/s per APA is sufficent to transfer it to the DAQ, with an order of magnitude safety factor. The link from the SP cryostat to the CUC will be implemented as either eight 1000Base-SX links or a single 10GBase-LR link per APA. The data on the links will be encoded using UDP/IP.

1.3.3 Offline Computing

The interface between the DAQ and offline computing is described in [5]. The DAQ team is responsible for reducing the data volume to the level that is agreed upon by all interested parties, and the raw data files are transfered from SURF to Fermilab using a dedicated network connection. A disk buffer is provided by the DAQ on or near the SURF site to hold several days worth of data so that the operation of the experiment is not affected if there happens to be a network disruption between SURF and Fermilab.

During stable running, the data volume produced by the DAQ systems of all four detector modules will be no larger than 30 PB/year. The maximum data rate is expected to be independent of the number of detector modules that are operational. During the construction of the second, third, and fourth detector modules, the extra rate per detector module will be used to gather data to aid in the refinement of the data selection algorithms. During commissioning, the data rate is expected to be higher than nominal running and it is anticipated that a data volume corresponding to (order) one year will be necessary to commission a detector module.

The disk buffer at SURF is planned to be 300 TB in size. The data link from SURF to Fermilab will support 100 Gbit/s (30 PB/year corresponds to about 8 Gbit/s). The offline computing team is responsible for developing the software to manage the transfer of files from SURF to Fermilab. The DAQ team is responsible for producing a reference implementation of the software that is used to access and decode the raw electronics data. The offline group is also responsible for providing the framework for real-time data quality monitoring (DQM). This monitoring is distinct from the online monitoring (OM). Developing the payload jobs that run various algorithms to summarize the data is the joint responsibility of the DAQ, offline, reconstruction and other groups. The DQM system includes a visualization system that can be accessed from the Internet and shows specifically where operation shifts are performed.

1.3.4 Slow Control

The cryogenic instrumentation and slow controls (CISC) systems monitor detector hardware and conditions not directly involved in taking the data described above. That data is stored both locally (in CISC database servers in the CUC) and offline (the databases will be replicated back to Fermilab) in a relational database indexed by timestamp. This allows bi-directional communications between the DAQ and CISC by reading or inserting data into the database as needed for non time-critical information.

For prompt, time sensitive status information such as *run is in progress* or *camera is on*, a low-latency software status register is available on the local network to both systems.

There is no hardware interface. However, several racks of CISC servers are in the counting room of the CUC, and rack monitors in DAQ racks are read out into the CISC data stream.

Note that life and hardware safety-critical items will be hardware interlocked according to Fermilab standards, and fall outside the scope of this interface.

1.3.5 External Systems

The DAQ is required to save data based on external triggers, e.g., when a pulse of beam neutrinos arrives at the FD; or upon notice of an interesting astrophysical event by SNEWS [6] or LIGO. This could involve going back to save data that has already been buffered (see Section 1.2.2), or changing the trigger or zero suppression criteria for data taken during the interesting time period.

1.3.5.1 Beam Trigger

The method for predicting and receiving the time of the beam spill is described in Section 1.2.6.1. Once that time is known to the DAQ, a high-level trigger can be issued to ensure that the necessary full data can be saved from the buffer and saved as an event.

1.3.5.2 Astrophysical Triggers

SuperNova Early Warning System (SNEWS) is a coincidence network of neutrino experiments that are individually sensitive to an SNB observed from a core-collapse supernova somewhere in our galaxy. While DUNE must be sensitive to such a burst on its own, and is expected to be able to contribute to the coincidence network (Section 1.2.5) via a TCP/IP socket, the capability to save data based on other observations provides an additional opportunity to ensure capture of this rare and valuable data. A SNEWS alert is formed when two or more neutrino experiments report a potential SNB signal within 10 s. A script running on the SNEWS server at BNL, provided by a given experiment that wishes to receive an alert, sends out a message with the earliest time in the

coincidence. The latency from the neutrino burst is set by the response time of the second fastest detector to report to SNEWS. This could be as short as seconds, but could be tens of seconds. At latencies larger than 10 s, full data might not be available, but selected data is expected to be manageable.

Other astrophysical triggers are available to which DUNE alone is unlikely to have sensitivity, except in rare cases, or if the triggers are taken as an ensemble. Among these are gravitational wave triggers (the details are being worked out during the current LIGO shutdown), and high-energy photon transients, most notably gamma ray bursts. In fact, the use of network sockets on the timescale of seconds enabled cooperation between LIGO, VIRGO, the Gamma Ray Coordinates Network (GCN) [1], and a number of automated telescopes to make the discovery that *short/hard* gamma ray bursts are caused by colliding neutron stars [7].

1.4 Production and Assembly

1.4.1 DAQ Components

The FD DAQ system comprises the classes of components listed below. In each case, we outline the production, procurement, quality assurance (QA), and quality control (QC) strategies.

1.4.1.1 Custom Electronic Modules

Custom electronic modules, specified and designed by the DAQ consortium, are used for two functional components in the DAQ FE. The first is to interface the detector module electronics to the DAQ FEC systems, which are likely to be based on the FELIX PCIe board. The other is for real-time data processing (particularly for the SP module), which will likely be based on the COB ATCA blade. ProtoDUNE-SP currently implements both designs, and new designs optimized according to DUNE requirements will be developed. It is possible that we will make use of commercially-designed hardware in one or other of these roles. DAQ consortium institutes have significant experience in the design and production of high-performance digital electronics for previous experiments. Our strategy is therefore to carry out design in-house, manufacturing and QA steps in industry, and testing and QC procedures at a number of specialized centers within the DUNE collaboration. Where technically and economically feasible, modules will be split into subassemblies (e.g., carrier board plus processing daughter cards), allowing production tasks to be spread over more consortium institutes.

DUNE electronic hardware will be of relatively high performance by commercial standards, and will contain high-value subassemblies such as large FPGAs. Achieving a high yield will require significant effort in design verification, prototyping and pre-production tests, as well as in tendering and vendor selection. The production schedule is largely driven by these stages and the need for

[1] Described in detail at `https://gcn.gsfc.nasa.gov/gcn_describe.html`

thorough testing and integration with firmware and software before installation, rather than by the time for series hardware manufacture. This is somewhat different from the majority of other DUNE FD components.

1.4.1.2 Commercial Computing

The majority of procured items will be standard commercial computing equipment, in the form of compute and storage servers. Here, the emphasis is on correct definition of the detailed specification, and the tracking of technology development, in order to obtain the best value during the tendering process. Computing hardware will be procured in several batches, as the need for DAQ throughput increases during the construction period.

1.4.1.3 Networking and data links

The data movement system is a combination of custom optical links (for data transmission from the cryostats to the CUC) and commercial networking equipment. The latter items will be procured in the same way as other computing components. The favored approach to procurement of custom optics is purchase of pre-manufactured assemblies ready for installation, rather than on-site fiber preparation and termination. Since transmission distances and latencies in the underground area are not critical, the fiber run lengths do not need to be of more than a few variants. It is assumed that fibers will not be easily accessible for servicing or replacement during the lifetime of the experiment, meaning that procurement and installation of spare *dark* fibers (including, if necessary, riser fibers up the SURF hoist shafts) is necessary.

1.4.1.4 Infrastructure

All DAQ components will be designed for installation in 48.3 cm (standard 19 in) rack infrastructure, either in the CUC or above ground. Standard commercial server racks with local air-water heat exchangers are likely to be used. These items will be specified and procured within the consortium, but will be pre-installed (along with the necessary electrical, cooling and safety infrastructure) under the control of technical coordination (TC) before DAQ beneficial occupancy.

1.4.1.5 Software and firmware

The majority of the DAQ construction effort will be invested in the production of custom software and firmware. Based on previous experiments, these projects are likely to use tens to hundreds of staff-years of effort, and will be significant projects even by commercial standards, mainly due to the specialized skills required for real-time software and firmware. A major project management effort is required to guide the specification, design, implementation and testing of the necessary

components, especially as developers will be distributed around the world. Use of common components and frameworks across all areas of the DAQ is mandatory. Effective DAQ software and firmware development has been a demonstrated weakness of several previous experiments, and substantial work is required in the next two years to put in place the necessary project management regime.

1.4.2 Quality Assurance and Quality Control

High availability is a basic requirement for the DAQ, and this rests upon three key principles:

- A rigorous QA and QC regime for components (including software and firmware);
- Redundancy in system design, to avoid single points of failure;
- Ease of component replacement or upgrade with minimal downtime.

The lifetime of most electronic assemblies or commercial computing components will not match the 20 year lifespan of the DUNE experiment. It is to be expected that essentially all components will therefore be replaced during this time. Careful system design will allow this to take place without changes to interfaces. However, it is intended that the system run for at least the first three to four years without substantial replacements, and QA and QC, as well as spares production, will be steered by this goal. Of particular importance is adequate burn-in of all components before installation underground, and careful record-keeping of both module and subcomponent provenance, in order to identify systematic lifetime issues during running.

1.4.3 Integration testing

Since the DAQ will use subcomponents produced by many different teams, integration testing is a key tool in ensuring compatibility and conformance to specification. This is particularly important in the prototyping phase before the design of final hardware. Once pre-production hardware is in hand, an extended integration phase will be necessary in order to perform final debugging and performance tuning of firmware and software. In order to facilitate ongoing optimization in parallel with operations, and compatibility testing of new hardware or software, we envisage the construction of one or more permanent integration test stands at DAQ institutions. These will be in locations convenient to the majority of consortium members, i.e., at major labs in Europe and the USA. A temporary DAQ integration and testing facility near SURF will also be required as part of the installation procedure.

1.5 Installation, Integration and Commissioning

1.5.1 Installation

The majority of DAQ components will be installed in a dedicated and partitioned area of the CUC as shown in Figure 1.6, starting as soon as the consortium has beneficial occupancy of this space. The conventional facilities (CF) is responsible for running fiber from the SP module's WIBs to the DAQ, and from the DAQ to the surface. This is currently projected to take place eighteen months before anode plane assemblies are installed in the SP module, allowing time for final component acceptance testing in the underground environment, and to prepare the DAQ for detector testing and commissioning. Some DAQ components (event builder, storage cluster and WAN routers, plus any post-event-builder processing) will be installed above ground.

A total of 500 kVA of power and cooling will be available to run the computers in the counting room. Twelve 48.3 cm (standard 19 in) server racks (of up to 58U height) per module have initially been allocated for each detector module, with two more each for facilities and CISC. An optimized layout, including the necessary space for hardware installation and maintenance, plus on-site spares, will be developed once the DAQ design is finalized. The racks will be water cooled with local air-to-water heat exchangers. To allow expanded headroom for initial testing, development, and commissioning throughput, the full complement of rack infrastructure and network equipment for four detector modules will be installed from the start.

The counting room is similar to a server room at a university or national lab in terms of the need for cleanliness, ventilation, fire protection, drop flooring, and access control. Networking infrastructure and fiber breakout will take up some of the rack space, but very little of the power budget. Power to individual machines and crates must to be controlled remotely via power distribution units, since it is desirable to minimize DAQ workers' presence underground if there is work that can be done from the surface or remotely. Some uninterruptible power supply (UPS) capacity is needed to allow for an orderly shutdown of computers, but only networking equipment requires longer-duration backup power, this is to enable remote recovery from short-term power failures.

1.5.2 Integration with Detector Electronics

Basic technical integration with detector electronics will take place before installation, during a number of integration exercises in the preceding years. We anticipate that the consortium will supply and support small-scale instances of the DAQ system for testing of readout hardware at the production sites, based on prototype or pre-production hardware. Full-scale DAQ testing will have been completed with artificial data sources during internal integration. The work to be done during installation is therefore essentially channel-by-channel verification of the final system as it is installed, on a schedule allowing for any rectifying work to be carried out on the detector immediately (i.e., the DAQ must gather and present data in effectively real time). This implies the presence of a minimal but sufficient functional DAQ system before detector installation commences, along with the timing and fast control system, and the capability to permanently record data for

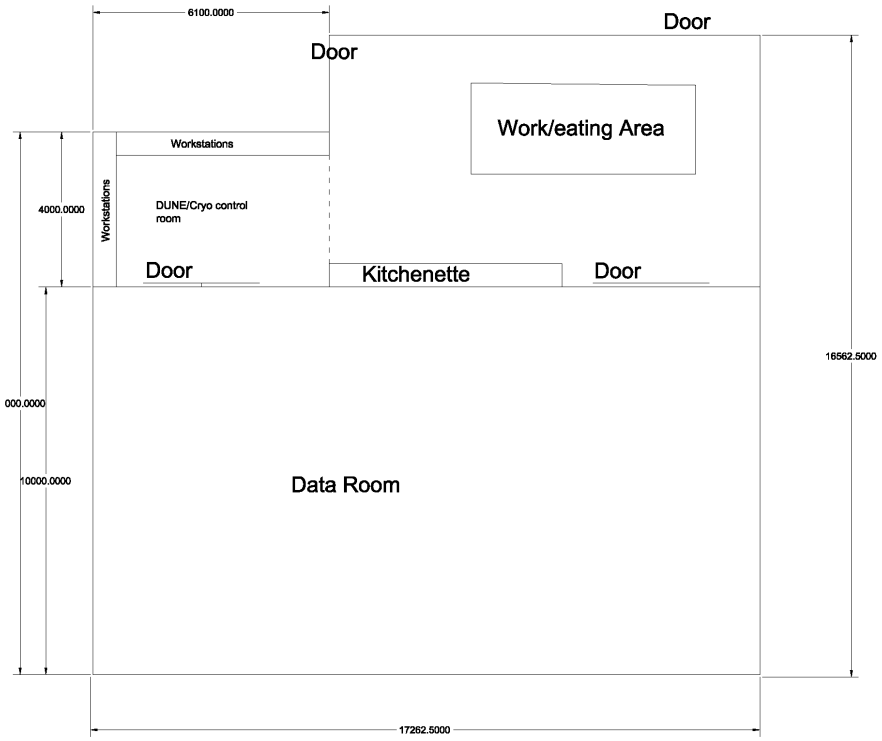

Figure 1.6: Floor plan for the DAQ and control room space in the CUC. The DAQ Room has space for at least 52 racks of servers and routers. Fiber from the WIBs in the detector caverns enter in the upper right of this room, terminate in a breakout panel, and are distributed to the RCEs in these racks, then to FELIX servers (also in this room) as outlined in Figure 1.3. Fibers to the surface enter this room from the lower left.

offline analysis. However, it does not require triggering, substantial event building or data transfer capacity. The DAQ installation schedule is essentially driven by this requirement.

In addition, the data pipeline from event builder, via the storage buffer and WAN, to the offline computing facilities, must be developed and tested. We anticipate this work largely happening at Fermilab in parallel with detector installation, and the full-scale instances of these components being installed at SURF in preparation for start of data-taking.

1.5.3 Commissioning

System commissioning for the DAQ comprises the following steps:

- Integration with detector subsystems of successive detector modules;
- Final integration and functional testing of all DAQ components;
- Establishment of the necessary tools and procedures to achieve high-efficiency operation;
- Selection, optimization and testing of trigger criteria;
- Ongoing and continuous self-test of the system to identify actual or imminent failures, and to assess performance.

Each of these steps will have been carried out at the integration test stands before being used on the final system. The final steps are to some extent continuous activities over the experiment lifetime, but which require knowledge of realistic detector working conditions before final validation of the system can take place. We anticipate that these steps will be carried out during the cryostat filling period, and form the major focus of the DAQ consortium effort during this time.

1.6 Safety

Two overall safety plans will be followed by the FD DAQ. General work underground will comply with all safety procedures in place for working in the detector caverns and CUC underground at SURF. DAQ-specific procedures for working with racks full of electronics or computers, as defined at Fermilab, will be followed, especially with respect to electrical safety and the fire suppression system chosen for the counting room. For example, a glass wall between the server room space and the other areas in Figure 1.6 will be necessary to prevent workers in the server room from being unseen if they are in distress, and an adequate hearing protection regime must be put in place.

There are no other special safety items for the DAQ system not already covered by the more general safety plans referenced above. The long-term emphasis is on remote operations capability from around the world, limiting the need for physical presence at SURF, and with underground

access required only for urgent interventions or hardware replacement.

1.7 Organization and Management

At the time of writing, the DAQ consortium comprises 30 institutions, including universities and national labs, from five countries. Since its conception, the DAQ consortium has met on roughly a weekly basis, and has so far held two international workshops dedicated to advancing the FD DAQ design. The current DAQ consortium leader is from U. Bristol, UK.

Several key technical and architectural decisions have been made in the last months, that have formed an agreed basis for the DAQ design and implementation presented in this document.

1.7.1 DAQ Consortium Organization

The DUNE DAQ consortium is currently organized in the form of five active Working Groups (WG) and WG leaders:

- Architecture, current WG leaders are from: U. Oxford and CERN;
- Hardware, current WG leaders are from: U. Bristol and SLAC;
- Data selection, current WG leader is from: U. Penn.;
- Back-end, current WG leader is from: Fermilab;
- Integration and Infrastructure, current WG leader is from: U. Minnesota Duluth.

During the ongoing early stages of the design, the architecture and hardware WGs have been holding additional meetings focused on aspects of the design related to architecture solutions and costing. In parallel, the DAQ Simulation Task Force effort, which was in place at the time of the consortium inception, has been adopted under the data selection WG, and simulation studies have continued to inform design considerations. This working structure is expected to remain in place through at least the completion of the interim design report (IDR). During the construction phase of the project we anticipate a new organization, built around major subsystem construction and commissioning responsibilities, and drawing also upon expertise build up during the ProtoDUNE projects.

1.7.2 Planning Assumptions

The DAQ planning is based the assumption of a SP module first, followed by a DP module. The schedule is sensitive to this assumption, as the DAQ requirements for the two module types are quite different. Five partially overlapping phases of activity are planned (see Figure 1.7):

- A further period of R&D activity, beginning at the time of writing, and culminating in a documented system design in the technical design report (TDR) around July 2019;
- Production and testing of a full prototype DAQ slice of realistic design, culminating in an engineering design review;
- Preparation and fit out of the CUC counting room with a minimal DAQ slice, in support of the first module installation;
- Production and delivery of final hardware, computing, software and firmware for the first module;
- Production and delivery of final hardware, computing, software and firmware for the second module.

This schedule assumes beneficial occupancy of the CUC ounting room by end of the first quarter of 2022, and the availability of facilities to support an extended large-scale integration test in 2020 (e.g., CERN or Fermilab). We assume the availability of resources for installation and commissioning of final DAQ hardware (e.g., surface control room and server room facilities) from around the first quarter of 2023, and the integration and test facility (ITF) from the second quarter of 2022. The majority of capital resources for DAQ construction will be required from the second quarter of 2022, with a first portion of funds for the minimal DAQ slice from the first quarter of 2021.

1.7.3 High-level Cost and Schedule

The high-level DAQ schedule, which is based upon the current DUNE FD top-level schedule, is shown in Figure 1.7.

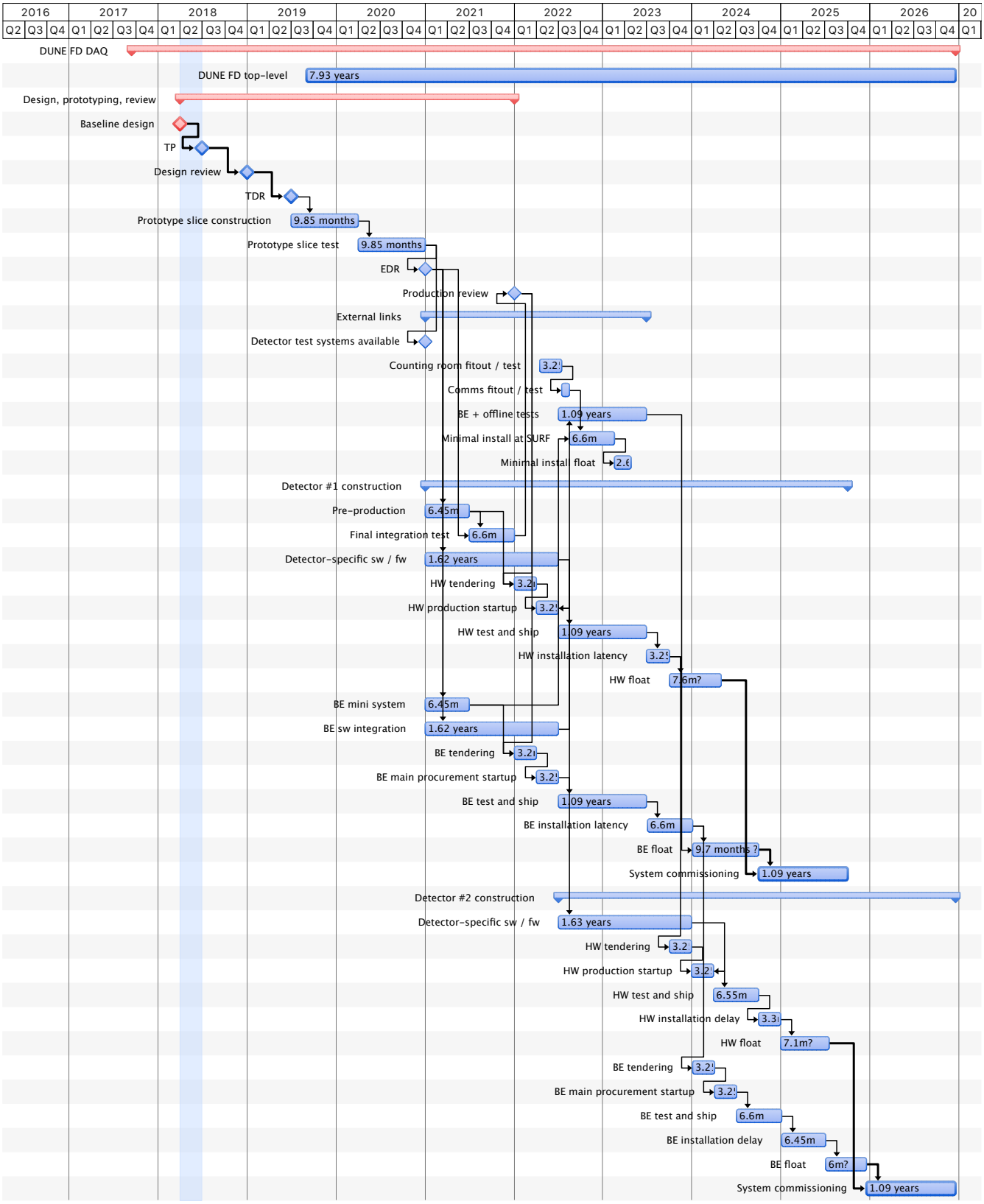

Figure 1.7: DAQ high-level schedule

Glossary

35 ton prototype The 35 ton prototype cryostat and single-phase (SP) detector built at Fermilab before the ProtoDUNE detectors. 9

analog-to-digital converter (ADC) A sampling of a voltage resulting in a discrete integer count corresponding in some way to the input. 17

advanced mezzanine card (AMC) Holds digitizing electronics and lives in Micro Telecommunications Computing Architecture (μTCA) crates. 26

anode plane assembly (APA) A unit of the SP detector module containing the elements sensitive to ionization in the LAr. It contains two faces each of three planes of wires, and interfaces to the cold electronics and photon detection system. 3, 9, 11, 13, 14, 17, 18, 26, 29, 30, 35, 43

***art*DAQ** A general-purpose Fermilab data acquisition framework for distributed data combination and processing. It is designed to exploit the parallelism that is possible with modern multicore and networked computers. 19

Advanced Telecommunications Computing Architecture (ATCA) An advanced computer architecture specification developed for the telecommunications, military, and aerospace industries that incorporates the latest trends high-speed interconnect technologies, next-generation processors, and improved reliability, availability and serviceability. 17, 29

Bump On Wire (BOW) A working name for the front-end readout computing elements used in the nominal DAQ design to interface the DP crates to the DAQ front-end computers. 13

cold electronics (CE) Refers to readout electronics that operate at cryogenic temperatures. 4, 9, 29

conventional facilities (CF) Pertaining to construction and operation of buildings or caverns and conventional infrastructure. 35

cryogenic instrumentation and slow controls (CISC) A DUNE consortium responsible for the cryogenic instrumentation and slow controls components. 31, 35

Cluster On Board (COB) An ATCA motherboard housing four RCEs. 14, 17, 18, 29, 32, 45

charge readout (CRO) The system for detecting ionization charge distributions in a DP detector module. 3, 9, 11, 13, 18, 29

central utility cavern (CUC) The central underground cavern containing utilities such as central cryogenics and other systems, and the underground data center and control room. iii, 11, 24–26, 29–31, 33, 35–37, 39

DAQ primary buffer (primary buffer) The portion of the DAQ front-end fragment that accepts full data stream from the corresponding detector unit and retains it sufficiently long for it to be available to produce a data fragment. 11, 13, 16, 18–20, 27, 42, 43

DAQ data receiver (DDR) The portion of the DAQ front-end fragment that accepts data from the DAQ front-end readout (FER), emits trigger candidates produced from the input trigger primitives, and forwards the full data stream to the DAQ primary buffer (primary buffer). 11, 13, 43

data selector The portion of the DAQ front-end fragment that accepts trigger commands and returns the corresponding data fragment. 11, 13, 16, 19, 43

DAQ front-end readout (FER) The portion of a DAQ front-end fragment that accepts data from the detector electronics and provides it to the front-end computer (FEC). In the nominal design it is also responsible for generating channel level trigger primitives. 7, 8, 13, 14, 16, 19, 21, 22, 24, 42, 43

DAQ front-end fragment The portion of one DAQ partition relating to a single FEC and corresponding to an integral number of detector units. See also data fragment. 4–7, 9, 12–14, 16–23, 29, 42

out-of-band trigger command dispatcher (OOB dispatcher) This component is responsible for dispatching a supernova neutrino burst (SNB) dump command to all FERs in the detector module. 7, 11, 19

DAQ partition A cohesive and coherent collection of DAQ hardware and software working together to trigger and read out some portion of one detector module; it consists of an integral number of DAQ front-end fragments. Multiple DAQ partitions may operate simultaneously, but each instance operates independently. 3–7, 13, 14, 21, 42, 43

data acquisition (DAQ) The data acquisition system accepts data from the detector FE electronics, buffers the data, performs a trigger decision, builds events from the selected data and delivers the result to the offline secondary DAQ buffer. i, iii, 2–9, 11–20, 22, 27–40, 43

data fragment A block of data read out from a single DAQ front-end fragment that span a contiguous period of time as requested by a trigger command. 7, 13, 16, 19, 22, 29, 42

detector module The entire DUNE far detector is segmented into four modules, each with a

nominal 10 kt fiducial mass. 2–4, 7–9, 12–14, 16–25, 28–30, 32, 35, 37, 42, 43, 45, 46

detector unit A subdetector may be partitioned into a number of similar parts. For example the single-phase TPC subdetector is made up of APA units. 3, 7, 12, 13, 20–22, 24, 29, 42–44, 47

secondary DAQ buffer A secondary DAQ buffer holds a small subset of the full rate as selected by a trigger command. This buffer also marks the interface with the DUNE Offline. 7, 8, 13, 18, 19, 42

DP module dual-phase detector module. 10, 11, 24, 39

dual-phase (DP) Distinguishes one of the DUNE far detector technologies by the fact that it operates using argon in both gas and liquid phases. 3, 13, 18, 21, 24, 25, 29

data quality monitoring (DQM) Analysis of the raw data to monitor the integrity of the data and the performance of the detectors and their electronics. This type of monitoring may be performed in real time, within the data acquisition (DAQ) system, or in later stages of processing, using disk files as input. 30

event builder (EB) A software agent servicing one detector module by executing trigger commands by reading out the requested data. 7, 8, 11, 13, 16, 18, 19

external trigger logic (ETL) Trigger processing that consumes detector module level trigger notification information and other global sources of trigger input and emits trigger command information back to the module trigger logics (MTLs). 7, 21–24, 44, 46

external trigger candidate Information provided to the MTL about events external to a detector module so that it may be considered in forming trigger commands. 22, 23, 44

far detector (FD) Refers to the 40 kt fiducial mass DUNE detector to be installed at the far site at SURF in Lead, SD, to be composed of four 10 kt modules. i, 2–6, 8, 31–33, 37–39

front-end computer (FEC) The portion of one DAQ partition that hosts the DAQ data receiver (DDR), primary buffer and data selector. It is connected to the FER via fiber optic. Each detector unit of a certain granularity, such as two SP anode plane assemblies (APAs), has one front-end computer that receives data from the readout hardware, hosts the primary DAQ memory buffer for that data, emits trigger candidates derived from that data, and satisfies requests for producing subsets of that data for egress. 4, 7–9, 12–14, 17, 18, 20–22, 32, 42

Front-End Link eXchange (FELIX) A high-throughput interface between front-end and trigger electronics and the standard PCIe computer bus. 14, 16, 18, 32, 36

front-end mother board (FEMB) Refers a unit of the SP cold electronics that contains the front-end amplifier and ADC ASICs covering 128 channels. 29

front-end (FE) The front-end refers a point that is "upstream" of the data flow for a particular

subsystem. For example the front-end electronics is where the cold electronics meet the sense wires of the TPC and the front-end DAQ is where the DAQ meets the output of the electronics. iii, 4–6, 8, 11, 12, 14, 17, 18, 20, 28, 29, 32, 46

field programmable gate array (FPGA) An integrated circuit technology that allows the hardware to be reconfigured to execute different algorithms after its manufacture and deployment. 14, 18, 32, 45

high-level trigger (HLT) A source of triggering at the module level. 21, 23

high voltage (HV) Generally describes a voltage applied to drive the motion of free electrons through some media. 4

integration and test facility (ITF) A facility where various detector components will be tested prior to installation. 39

liquid argon time-projection chamber (LArTPC) A class of detector technology that forms the basis for the DUNE far detector modules. It typically entails observation of ionization activity by electrical signals and of scintillation by optical signals. 2

long-baseline (LBL) Refers to the distance between the neutrino source and the far detector. It can also refer to the distance between the near and far detectors. The "long" designation is an approximate and relative distinction. For DUNE, this distance (between Fermilab and SURF) is approximately 1300 km. 3

light readout (LRO) The system for detecting scintillation photons in a DP detector module. 3, 13, 18

module trigger logic (MTL) Trigger processing that consumes detector unit level trigger command information and emits trigger commands. It provides the external trigger logic (ETL) with trigger notifications and receives back any external trigger candidates. 8, 11, 13, 14, 16–24, 43, 46

online monitoring (OM) Processes that run inside the DAQ on data "in flight," specifically before landing on the offline disk buffer, and that provide feedback on the operation of the DAQ itself and the general health of the data it is marshalling. 11, 30

ProtoDUNE-SP The SP ProtoDUNE detector. 24, 25, 32

photon detection system (PDS) The detector subsystem sensitive to light produced in the LAr. 2, 7, 24, 30

photon detector (PD) Refers to the detector elements involved in measurement of number and arrival times of optical photons produced in a detector module. 4, 12

one-pulse-per-second signal (1PPS signal) An electrical signal with a fast rise time and that arrives in real time with a precise period of one second. 24, 26

ProtoDUNE Either of the two DUNE prototype detectors constructed at CERN and operated in a CERN test beam (expected fall 2018). One prototype implements SP and the other DP technology. 2, 19, 38, 41

quality assurance (QA) The process by which quality is maintained so as to preserve high availability and precise function. 32, 34

quality control (QC) A system of maintaining quality through testing products against a specification. 32, 34

DAQ event block The unit of data output by the DAQ. It contains trigger and detector data spanning a unique, contiguous time period and a subset of the detector channels.. 7, 13, 19, 20

reconfigurable computing element (RCE) Data processor located outside of the cryostat on a Cluster On Board (COB) which contains field programmable gate array (FPGA), RAM and SSD resources, responsible for buffering data, producing trigger primitives, responding to triggered requests for data and sinking SNB dumps. 13, 14, 16–18, 26, 36

run control (RC) The system for configuring, starting and terminating the DAQ. 23

readout window A fixed, atomic and continuous period of time over which data from a detector module, in whole or in part, is recorded. This period may differ based on the trigger that initiated the readout. 21, 23, 24

radio frequency (RF) Electromagnetic emissions that are within the (radio) frequency band of sensitivity of the detector electronics. 18

S/N signal-to-noise (ratio). 30

SBN Short-Baseline Neutrino program (at Fermilab). 8

silicon photomultiplier (SiPM) A solid-state avalanche photodiode sensitive to single photoelectron signals. 30

spill location system (SLS) A system residing at the DUNE far detector site that provides information, possibly predictive, indicating periods of time when neutrinos are being produced by the Fermilab Main Injector beam spills. 27

supernova neutrino burst (SNB) A prompt increase in the flux of low-energy neutrinos emitted in the first few seconds of a core-collapse supernova. It can also refer to a trigger command type that may be due to an SNB, or detector conditions that mimic its interaction signature. 2–8, 10, 12–16, 18–24, 31, 42, 45

SuperNova Early Warning System (SNEWS) A global supernova neutrino burst trigger formed by a coincidence of SNB triggers collected from participating experiments. 21, 23, 31, 32

SP module single-phase detector module. i, 9, 10, 16, 20, 21, 24, 29, 32, 35, 39

single-phase (SP) Distinguishes one of the DUNE far detector technologies by the fact that it operates using argon in its liquid phase only . 3, 13, 14, 16, 18, 20, 21, 24, 25, 29, 30, 41, 43

solid-state disk (SSD) Any storage device that may provide sufficient write throughput to receive, both collectively and distributed, the sustained full rate of data from a detector module for many seconds. 8, 13, 14, 16, 19, 20

subdetector A detector unit of granularity less than one detector module such as the TPC of either a SP or DP module. 12, 43

technical coordination (TC) A dedicated DUNE project effort responsible for assuring proper interfaces between the collaboration consortia. 33

technical design report (TDR) A formal project document that describes the experiment at a technical level. 39, 46

time projection chamber (TPC) The portion of each DUNE detector module that records ionization electrons after they drift away from a cathode through the LAr, and also through gaseous argon in a DP module. The activity is recorded by digitizing the waveforms of current induced on the anode as the distribution of ionization charge passes by or is collected on the electrode. 2

interim design report (IDR) An intermediate milestone on the path to a full technical design report (TDR). 38

trigger candidate Summary information derived from the full data stream and representing a contribution toward forming a trigger decision. 7, 13, 14, 16–24, 46

trigger command Information derived from one or more trigger candidates that directs elements of the detector module to read out a portion of the data stream. 7, 13, 16, 19–21, 23, 24, 42–44, 46

trigger decision The process by which trigger candidates are converted into trigger commands. 7, 9, 12, 16, 28, 42, 46

trigger notification Information provided by MTL to ETL about trigger decision its processing. 43, 44

trigger primitive Information derived by the DAQ front-end (FE) hardware that describes a region of space (e.g., one or several neighboring channels) and time (e.g., a contiguous set of ADC sample ticks) associated with some activity. 4, 7, 11, 13, 14, 17–22, 42

Micro Telecommunications Computing Architecture (μTCA) The computer architecture specification followed by the crates that house charge and light readout electronics in the dual-phase module. 24–26, 41

warm interface board (WIB) Digital electronics situated just outside the SP cryostat that receives digital data from the FEMBs over cold copper connections and sends it to the RCE FE readout hardware. 3, 14, 16, 17, 24, 29, 35, 36

White Rabbit (WR) A component of the timing system that forwards clock signal and time-of-day reference data to the master timing unit. 24, 25

zero-suppression (ZS) Used to delete some portion of a data stream that does not significantly deviate from zero or intrinsic noise levels. It may be applied at different granularity from per-channel to per detector unit. 22

References

[1] DOE Office of High Energy Physics, "Mission Need Statement for a Long-Baseline Neutrino Experiment (LBNE)," tech. rep., DOE, 2009. LBNE-doc-6259.

[2] D. Newbold, "ProtoDUNE-SP Timing System: Interfaces and Protocol," tech. rep., Bristol, 2016. http://docs.dunescience.org/cgi-bin/ShowDocument?docid=1651.

[3] D. Christian, G. Karagiorgi, D. Newbold, and M. Verzocchi, "DUNE FD Interface Document: SP TPC Electronics to joint DAQ," tech. rep., Fermilab, Columbia, and Bristol, 2018. http://docs.dunescience.org/cgi-bin/ShowDocument?docid=6742.

[4] D. Duchesneau, I. Gil-Botella, G. Karagiorgi, and D. Newbold, "DUNE FD Interface Document: SP Photon Detector to Joint DAQ," tech. rep., LAPP, CIEMAT, Columbia, and Bristol, 2018. http://docs.dunescience.org/cgi-bin/ShowDocument?docid=6727.

[5] G. Karagiorgi, D. Newbold, A. Norman, and H. Schellman, "DUNE FD Interface Document: Software and Computing to Joint DAQ," tech. rep., Columbia, Bristol, Fermilab, and Oregon State, 2018. http://docs.dunescience.org/cgi-bin/ShowDocument?docid=7123.

[6] P. Antonioli *et al.*, "SNEWS: The SuperNova Early Warning System," *New J. Phys.* **6** (2004) 114, arXiv:astro-ph/0406214.

[7] **GROND, SALT Group, OzGrav, DFN, INTEGRAL, Virgo, Insight-Hxmt, MAXI Team, Fermi-LAT, J-GEM, RATIR, IceCube, CAASTRO, LWA, ePESSTO, GRAWITA, RIMAS, SKA South Africa/MeerKAT, H.E.S.S., 1M2H Team, IKI-GW Follow-up, Fermi GBM, Pi of Sky, DWF (Deeper Wider Faster Program), Dark Energy Survey, MASTER, AstroSat Cadmium Zinc Telluride Imager Team, Swift, Pierre Auger, ASKAP, VINROUGE, JAGWAR, Chandra Team at McGill University, TTU-NRAO, GROWTH, AGILE Team, MWA, ATCA, AST3, TOROS, Pan-STARRS, NuSTAR, ATLAS Telescopes, BOOTES, CaltechNRAO, LIGO Scientific, High Time Resolution Universe Survey, Nordic Optical Telescope, Las Cumbres Observatory Group, TZAC Consortium, LOFAR, IPN, DLT40, Texas Tech University, HAWC, ANTARES, KU, Dark Energy Camera GW-EM, CALET, Euro VLBI Team, ALMA** Collaboration, B. P. Abbott *et al.*, "Multi-messenger Observations of a Binary Neutron Star Merger," *Astrophys. J.* **848** no. 2, (2017) L12, arXiv:1710.05833 [astro-ph.HE].